

DISCOVERY AND DEVELOPMENT OF SMALL-
MOLECULE MODULATORS FOR THE SULFATION OF
GLYCOSAMINOGLYCANS
and
STUDYING THE ROLE OF *O*-GLCNAC ON CREB
THROUGH SEMISYNTHESIS

Thesis by
Sheldon Ting Fong Cheung

In Partial Fulfillment of the Requirements for
the degree of
Doctor of Philosophy

CALIFORNIA INSTITUTE OF TECHNOLOGY
Pasadena, California

2017
(Defended September 22, 2016)

© 2017

Sheldon T. Cheung
All Rights Reserved

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my advisor, Prof. Linda Hsieh-Wilson, for her support and guidance throughout the whole process. Without Linda, I would not have been able to work on such amazing research and accomplish the work reported here. She challenged me to be the best scientist possible and I would not be the scientist I am today without her direction. I would also like to thank my committee members, Prof. Bil Clemons, Prof. Peter Dervan, and Prof. Dennis Dougherty. Their scientific advice and support throughout my time here have also been instrumental to my success. Additionally, I would like to thank all Hsieh-Wilson lab members, past and present. I have been lucky to have such great colleagues to work with every day that can give day-to-day help and make the graduate school process that much more enjoyable. I would also like to thank my family: my mom, dad, brother Sab, sister Liv, mother-in-law Lai-Yoong, father-in-law Chee-Pun, and sister-in-law Xiao-Xiang. They have all not only been supportive of my studies 3,000 miles away, but have been part of a support network for all parts of my life as well. Last, but certainly not least, I would like to thank my wife Xiao-Lan and son Lewis. They are everything in the world to me and their loving support was what truly allowed me to get through this trying process.

ABSTRACT

Glycosaminoglycans (GAGs) are sulfated polysaccharides that play key roles in many cellular processes, ranging from viral invasion and cancer metastasis to neuronal development. Their diverse biological activities stem from their complex sulfation patterns, which are tightly regulated *in vivo*. For instance, the GAG chondroitin sulfate (CS) has been shown to undergo regiochemical sulfation during development and after spinal cord injury. However, few tools exist to modulate specific GAG sulfation patterns and study their importance in different biological contexts. Here, we identified the first cell-permeable small molecule that can selectively inhibit GAG sulfotransferases and modify the fine structure of GAGs. We demonstrate that the inhibitor reduces GAG sulfation *in vitro* and in cells and reverses CS-E-mediated inhibition of neuronal outgrowth. This small molecule may serve as a useful lead compound or chemical tool for studying the importance of CS and other GAGs in normal biology and disease.

The β -*N*-acetyl-D-glucosamine (*O*-GlcNAc) post-translational modification plays a major role in many diseases such as cancer, diabetes, and neurodegenerative disorders, but much is still unknown about its molecular-level influence on protein structure and function. Although post-translational modifications have been known to induce important structural changes in proteins, notably, no structures of *O*-GlcNAcylated proteins exist. The challenge of obtaining homogeneous glycoproteins bearing the GlcNAc sugar at defined sites has hindered the structural and biochemical studies of this modification. Here we have utilized a semisynthetic approach to generate a homogeneously *O*-GlcNAcylated form of cyclic-AMP response element binding protein (CREB) for structural and functional studies.

TABLE OF CONTENTS

| | |
|--|-----|
| Acknowledgements | iii |
| Abstract | iv |
| Table of Contents | v |
| List of Figures, Tables and/or Schemes | vi |
| List of Abbreviations | ix |
| Chapter 1: Introduction Glycosaminoglycans and the Role of Sulfation | 1 |
| References | 14 |
| Chapter 2: Chst15-Antagonist High-Throughput Screen and Validation | 18 |
| Introduction | 19 |
| Results and Discussion | 19 |
| Conclusion | 27 |
| Experimental Methods | 28 |
| References | 32 |
| Chapter 3: Structure-Activity Relationship Analysis of Lead Compounds | 34 |
| Introduction | 35 |
| Results and Discussion | 36 |
| Conclusion | 41 |
| Experimental Methods | 42 |
| References | 69 |
| Chapter 4: Mechanistic and Functional Characterization of Compound 37 | 70 |
| Introduction | 71 |
| Results and Discussion | 72 |
| Conclusion | 86 |
| Experimental Methods | 87 |
| References | 94 |
| Chapter 5: Introduction to <i>O</i> -GlcNAc and Studying its Role in Structure and Function | 96 |
| References | 105 |
| Chapter 6: Examining the Role of <i>O</i> -GlcNAc on CREB through Semisynthesis | 109 |
| Introduction | 110 |
| Results and Discussion | 110 |
| Conclusion | 119 |
| Experimental Methods | 120 |
| References | 132 |
| Appendix A: Synthesis of Uridine Diphosphate 6-Azido D-Galactose | 134 |
| Experimental Methods | 135 |

LIST OF FIGURES, TABLES, AND/OR SCHEMES

Chapter 1: Introduction to Glycosaminoglycans and the Role of Sulfation

| | |
|---|---|
| Figure 1.1 Structures of the major classes of sulfated glycosaminoglycans..... | 3 |
| Figure 1.2 Four major sulfation motifs of chondroitin sulfate | 5 |
| Figure 1.3 Illustrative representation of chondroitin sulfate biosynthesis | 9 |

Chapter 2: Chst15-Antagonist High-Throughput Screen and Validation

| | |
|---|----|
| Figure 2.1 High-throughput assay for the discovery of Chst15 small molecule inhibitors | 20 |
| Figure 2.2 Radioactive isotope labeling assay and Chst15 mutational effects on enzyme activity | 21 |
| Figure 2.3 Optimization of high-throughput assay by fluorimetry | 23 |
| Figure 2.4 PAPS and CS-A substrate K_m determination by fluorimetry | 24 |
| Figure 2.5 Identity of small molecule hits from the fluorescent high-throughput screen ... | 25 |
| Figure 2.6 Validation of hit compounds via the fluorescence assay | 26 |
| Figure 2.7 Validation of candidate compounds by radioactive isotope labeling assay..... | 27 |

Chapter 3: Structure-Activity Relationship Analysis of Lead Compounds

| | |
|---|----|
| Figure 3.1 Comparison of chemical analogs to lead compound 5 (Part I)..... | 37 |
| Scheme 3.1 Synthesis of compound 19 analogs (Route 1)..... | 38 |
| Scheme 3.2 Synthesis of compound 19 analogs (Route 2)..... | 38 |
| Table 3.1 Chst15 Inhibition Relative to 19 of Anilide-Analogs at 25 μ M..... | 40 |
| Table 3.2 Chst15 Inhibition Relative to 19 of Vinylbenzene-Analogs at 25 μ M..... | 40 |
| Figure 3.2 Comparison of additional chemical analogs to lead compound 19 | 41 |

Chapter 4: Mechanistic and Functional Characterization of Compound 37

| | |
|---|----|
| Figure 4.1 Compound 37 selectively inhibits glycosaminoglycan sulfotransferases. | 73 |
| Figure 4.2 Steady-state kinetic analysis reveals competitive and mixed modes of Chst15 inhibition by 37 | 74 |
| Table 4.1 Kinetic parameters of inhibition for Chst15 | 75 |
| Figure 4.3 Compound 37 exhibits reversible-covalent inhibitor properties when using β ME as a model nucleophile | 76 |
| Figure 4.4 Compound 37 inhibitory activity depends on pre-incubation time and shows reversibility | 77 |
| Figure 4.5 Representation of astrogliosis after neuronal injury and the potential therapeutic mechanism of small molecule treatment | 78 |
| Figure 4.6 Treatment of NIH/3T3 fibroblasts and Neu7 astrocytes with 37 leads to a significant, dose-dependent decrease in CS-E expression | 79 |
| Figure 4.7 Treatment of NIH/3T3 fibroblasts with 37 leads to a significant, dose-dependent decrease in HS expression | 80 |
| Figure 4.8 CSPGs from Neu7 astrocytes treated with inhibitor 37 show decreased levels of CS-E and other sulfation motifs | 81 |
| Table 4.2 Quantification of the percent disaccharide composition of CSPGs secreted from untreated and treated astrocytes | 81 |
| Figure 4.9 DRG treatment with 37 leads to no significant effect in neurite outgrowth | 82 |
| Figure 4.10 The inhibitory effects of CSPGs secreted from Neu7 astrocytes are attenuated by treatment of the cells with 37 | 83 |
| Figure 4.11 CSPGs from astrocytes treated with 37 repel fewer neurites compared to CSPGs from untreated astrocytes..... | 84 |
| Table 4.3 <i>In vitro</i> microsomal stability of 19 and 37 | 85 |
| Table 4.4 <i>In vitro</i> cytochrome p450 inhibition by 196 | 85 |
| Table 4.5 Pharmacokinetic parameters of 37 in rats | 86 |

Chapter 5: Introduction to *O*-GlcNAc and Studying its Role in Structure and Function

| | |
|---|-----|
| Figure 5.1 Dynamic, covalent attachment and removal of <i>O</i> -GlcNAc is catalyzed by OGT and OGA | 98 |
| Figure 5.2 Four domains of CREB and locations of key phosphorylation and <i>O</i> -GlcNAcylation sites | 101 |
| Figure 5.3 One-pot chemical glycosylation reaction through the conversion of dehydroalanine (Dha)..... | 105 |

Chapter 6: Examining the Role of *O*-GlcNAcylation on CREB through Semisynthesis

| | |
|--|-----|
| Figure 6.1 CREB mutant purification to >95% homogeneity..... | 111 |
| Figure 6.2 Functional characterization of WT and mutant CREB using different purification protocols..... | 113 |
| Figure 6.3 One-pot chemical glycosylation reaction of CREB..... | 114 |
| Table 6.1 Cys to <i>S</i> -GlcNAc Conversion with DBHDA | 114 |
| Figure 6.4 Intact protein LC-MS showed the rapid conversion of mutant CREB to Dha-containing CREB and to <i>S</i> -GlcNAc-containing CREB | 116 |
| Figure 6.5 CID-MS analysis of in-gel chymotryptic digest after bioconjugation reactions identified CREB peptides bearing Dha and <i>S</i> -GlcNAc at S40C position..... | 117 |
| Figure 6.6 Wheat germ agglutinin (WGA) purification of semi-synthetic CREB | 118 |
| Figure 6.7 CREB DNA binding was unaffected by S40C modifications..... | 119 |

LIST OF ABBREVIATIONS

| | |
|-------------------|---|
| AcOH | acetic acid |
| ADME | adsorption, distribution, metabolism, and elimination |
| AMAC | 2-aminoacridone |
| APCI | atmospheric pressure chemical ionization |
| APS | adenosine-5'-phosphosulfate |
| ATP | adenosine triphosphate |
| AUC | area under the curve |
| BME | 2-mercaptoethanol |
| BSA | bovine serum albumin |
| bZIP | basic leucine zipper domain |
| CaCl ₂ | calcium chloride |
| cAMP | cyclic-AMP |
| CBP | CREB binding protein |
| CCl ₄ | carbon tetrachloride |
| ChABC | chondroitinase ABC |
| Chst3 | chondroitin 6- <i>O</i> -sulfotransferase |
| Chst11 | chondroitin 4- <i>O</i> -sulfotransferase |
| Chst15 | <i>N</i> -acetylgalactosamine 4-sulfate 6- <i>O</i> -sulfotransferase |
| CID-MS | collision-induced mass spectrometry |
| CK2 | casein kinase 2 |
| CL | plasma clearance |
| C _{max} | maximum plasma concentration |
| CNS | central nervous system |
| CO ₂ | carbon dioxide |
| CRE | cyclic-AMP response element |
| CREB | cyclic-AMP response element binding protein |
| CRTC | CREB regulated transcription coactivator |
| CS | chondroitin sulfate |
| CSPG | chondroitin sulfate proteoglycan |
| DAPI | 4',6-diamidino-2-phenylindole |
| DBHDA | 2,5-dibromohexanediamide |
| DCM | dichloromethane |
| DEAE | diethylaminoethyl |
| Dha | dehydroalanine |
| DIC | <i>N,N'</i> -diisopropylcarbodiimide |
| DIPEA | <i>N,N</i> -diisopropylethylamine |
| DMEM | Dulbecco's modified Eagle media |
| DMF | dimethylformamide |
| DMSO | dimethyl sulfoxide |
| DNA | deoxyribose nucleic acid |

| | |
|---------------------------------|---|
| DRG | dorsal root ganglion |
| DS | dermatan sulfate |
| DTT | dithiothreitol |
| EDTA | ethylenediaminetetraacetic acid |
| EMSA | electrophoretic mobility shift assay |
| ESI | electrospray ionization |
| EtN ₃ | triethylamine |
| Et ₂ O | diethyl ether |
| EtOAc | ethyl acetate |
| FBS | fetal bovine serum |
| GAG | glycosaminoglycan |
| Gal | galactose |
| GalNAc | <i>N</i> -acetylgalactosamine |
| GlcA | glucuronic acid |
| GlcNAc | <i>N</i> -acetylglucosamine |
| GRIF-1 | GABA _A receptor-associated protein |
| HBr | hydrobromic acid |
| HCl | hydrochloric acid |
| HClO ₄ | perchloric acid |
| HDAC | histone deacetylase |
| Hex | hexanes |
| HPLC | high performance liquid chromatography |
| HRMS | high resolution mass spectra |
| HS | heparan sulfate |
| Hs3st1 | heparan 3- <i>O</i> -sulfotransferase 1 |
| HTS | high-throughput screen |
| IdoA | iduronic acid |
| IPTG | isopropyl β-D-1-thiogalactopyranoside |
| KCl | potassium chloride |
| KH ₂ PO ₄ | potassium phosphate monobasic |
| KID | kinase inducible domain |
| KIX | KID interacting domain |
| KS | keratan sulfate |
| LB | lysogeny broth |
| LC-MS | liquid chromatography-mass spectrometry |
| LTQ-FT | linear trap quadrupole-Fourier transform |
| Man | mannose |
| MBP | maltose binding protein |
| MeCN | acetonitrile |
| MeOH | methanol |
| MgCl ₂ | magnesium chloride |
| MgSO ₄ | magnesium sulfate |
| MsCl | methanesulfonyl chloride |
| MSH | <i>O</i> -mesitylenesulfonylhydroxylamine |
| MU | 4-methylumbelliferone |

| | |
|---|---|
| MUS | 4-methylumbelliferyl sulfate |
| MWCO | molecular weight cutoff |
| NaBH ₃ CN | sodium cyanoborohydride |
| NaCl | sodium chloride |
| NaHCO ₃ | sodium bicarbonate |
| Na ₂ HPO ₄ | sodium phosphate dibasic |
| NaH ₂ PO ₄ | sodium phosphate monobasic |
| NaOAc | sodium acetate |
| NaOH | sodium hydroxide |
| NaOMe | sodium methoxide |
| NaP _i | sodium phosphate |
| Na ₂ S ₂ O ₅ | sodium metabisulfite |
| NaSO ₄ | sodium sulfate |
| NBS | <i>N</i> -bromosuccinimide |
| NH ₄ HCO ₃ | ammonium bicarbonate |
| NH ₄ OAc | ammonium acetate |
| NH ₄ OH | ammonium hydroxide |
| Ni-NTA | nickel-nitrilotriacetic acid |
| OIP106 | <i>O</i> -GlcNAc transferase interacting protein 106 |
| OGA | <i>O</i> -GlcNAcase |
| <i>O</i> -GlcNAc | <i>O</i> -linked β-N-acetylglucosamine |
| OGT | <i>O</i> -GlcNAc transferase |
| PAP | 3'-phosphoadenosine-5'-phosphate |
| PAPS | 3'-phosphoadenosine-5'-phosphosulfate |
| PBS | phosphate buffered saline |
| PCR | polymerase chain reaction |
| Pd/C | palladium on carbon |
| PEP | phosphoenolpyruvic acid |
| PFA | paraformaldehyde |
| PhMe | toluene |
| PK | pharmacokinetic |
| Q1 | glutamine rich domain 1 |
| Q2 | glutamine rich domain 2 |
| rATP | deoxyadenosine triphosphate |
| rCTP | deoxycytidine triphosphate |
| rGTP | deoxyguanosine triphosphate |
| ROESY | rotating frame nuclear Overhauser effect spectroscopy |
| RPIP-HPLC | reversed-phase ion-pair HPLC |
| RT | room temperature |
| rUTP | deoxyuridine triphosphate |
| SDS-PAGE | sodium dodecyl sulfate polyacrylamide gel electrophoresis |
| SEC | size exclusion chromatography |
| <i>S</i> -GlcNAc | 1-thio-2-acetamido-2-deoxy-β-D-glucopyranose |
| SOCl ₂ | thionyl chloride |
| SNR | signal-to-noise ratio |

| | |
|------------------|---|
| Sult1c1 | aryl sulfotransferase IV |
| Sult1e1 | estrogen sulfotransferase |
| Sult2b1a | pregnenolone sulfotransferase |
| Sult2b1b | cholesterol sulfotransferase |
| T _{1/2} | half life |
| TAE | Tris/acetate/EDTA |
| TBE | Tris/borate/EDTA |
| t-BuOOH | <i>tert</i> -butyl hydroperoxide |
| THF | tetrahydrofuran |
| T _{max} | time of maximum plasma concentration |
| TPR | tetratricopeptide repeats |
| uCS | unsulfated chondroitin |
| UMP-morpholidate | 4'-morpholine- <i>N,N'</i> -dicyclohexylcarboxamidinium |
| Ust | uronyl sulfotransferase |
| V _{ss} | volume of distribution |
| WGA | wheat germ agglutinin |
| WT | wild type |
| Xyl | xylose |
| Z' | Z-factor |

Chapter 1

INTRODUCTION TO GLYCOSAMINOGLYCANS AND THE ROLE OF SULFATION

INTRODUCTION TO GLYCOSAMINOGLYCANS AND THE ROLE OF SULFATION

Glycosaminoglycans (GAGs) are a large class of polysaccharides that decorate cell surface proteins or exist as part of the extracellular matrix. They play a myriad of roles in different cellular processes, including viral invasion,^{1,2} cancer metastasis,³⁻⁵ neuronal development,^{6,7} and spinal cord injury.^{8,9} Their diverse biological activities stem from their complex and heterogeneous structures, which allow them to interact with a wide range of proteins. GAGs are long, linear polysaccharide chains composed of repeating disaccharide units that display various patterns of sulfation. Their length can vary from 20-200 disaccharide units and contain areas of low to high sulfation. It has been established that specific GAG sulfation motifs are key structural determinants in regulating protein recognition and downstream signaling. However, while nature chose to exploit GAG heterogeneities to mediate distinct cellular functions, these same heterogeneities have complicated our ability to study their specific biological roles. The study of GAGs would greatly benefit from the development of new tools, particularly small-molecule modulators, to help dissect the molecular properties that elicit these different biological functions.

Types of Glycosaminoglycans

There are three major sulfated classes of GAGs: heparin and heparan sulfate (HS), chondroitin sulfate (CS) and dermatan sulfate (DS), and keratan sulfate (KS). These three

families of GAGs have various biological functions despite only subtle differences in their macromolecular structures (Figure 1.1). All of these GAG structures are found linked to a proteoglycan core bound to the cell membrane or secreted into the extracellular matrix predominantly via *O*-linkages to serine or threonine residues. KS is the one exception that may also be found linked via *N*-linkages to asparagine residues.¹⁰⁻¹²

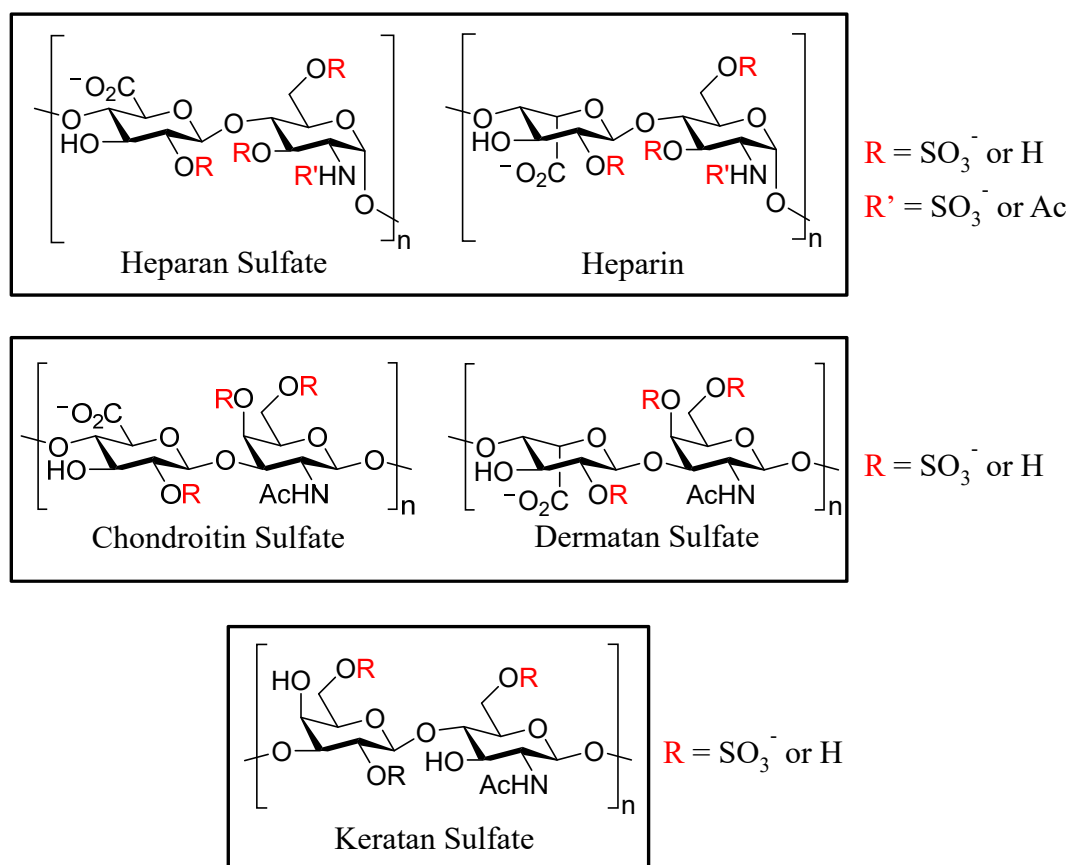


Figure 1.1. Structures of the major classes of sulfated glycosaminoglycans.

Heparin and HS, the most well studied class of GAGs, are particularly important in anticoagulation,¹³ cell growth and development,¹⁴ angiogenesis,¹⁵ and viral invasion.^{1,2} They are made up of disaccharide units containing *N*-acetylglucosamine (GlcNAc) and glucuronic

acid (GlcA) or iduronic acid (IdoA). Despite sharing the same structural components, the classification between heparin and HS is derived from their overall composition. Heparin, produced only by mast cells, exists as a shorter polysaccharide (10-12 kDa) compared to the ubiquitously expressed HS (10-70 kDa). Additionally, while the presence of IdoA predominates in heparin (>70%), GlcA represents the majority of uronic acid content in HS (50-70%). Another important distinction comes from the sulfation of the two types of molecules. HS contains an average of ≥ 1 sulfate groups per disaccharide, while heparin contains about ~ 2.7 sulfate groups.¹² Due to the amount of sulfate groups, heparin possesses the highest negative charge density of any known biological macromolecule. Heparin and HS also share the same positions in which sulfation or acetylation can occur: sulfation at the 3-*O* and 6-*O* of GlcNAc, sulfation of the 2-*O* of GlcA or IdoA, and *N*-sulfation or *N*-acetylation. Comparative analysis between heparin and HS has shown that the *N*-sulfation frequency is much higher (>80%) and the concentration of *O*-sulfated groups exceed that of *N*-sulfated groups in heparin.¹⁶

The next family of GAGs, chondroitin sulfate and dermatan sulfate, have intriguing functions in numerous cellular processes, including neuronal development,⁶ spinal cord injury,⁸ growth factor signaling,¹⁷ morphogenesis,¹⁸ and cell division.¹⁹ CS is comprised of disaccharide units containing GlcA and *N*-acetylgalactosamine (GalNAc), while DS disaccharide units are made up of IdoA and GalNAc. Despite having different uronic acid content, these two GAGs are classified in the same family because they share a similar biosynthetic pathway and linear polysaccharides often exist as co-polymers of both structures.¹¹ Sulfation is specifically localized to three hydroxyl groups, the 2-*O* of the uronic acid and the 4-*O* and 6-*O* of GalNAc. Defined combinations of these sulfate groups give rise

to the four major sulfation patterns: CS/DS-A, -C, -D, and -E (Figure 1.2). Expression of these different motifs often changes throughout development and disease and has been shown to molecularly facilitate different binding surfaces for interacting proteins.^{20,21}

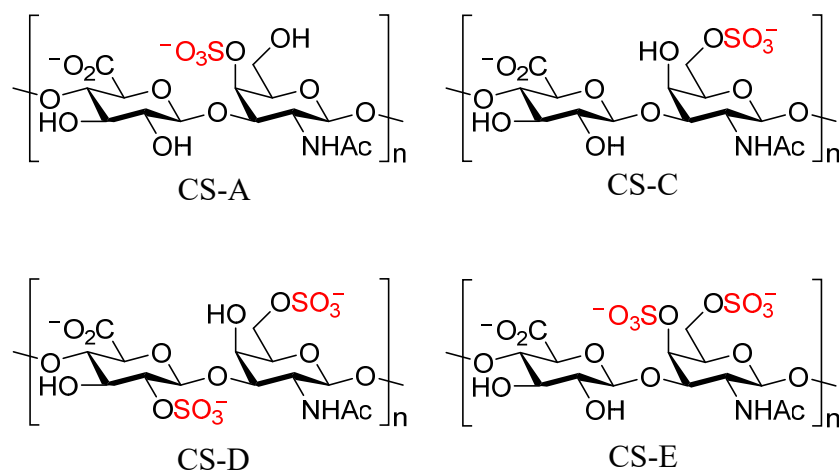


Figure 1.2. Four major sulfation motifs of chondroitin sulfate (CS).

KS, the last type of sulfated GAGs, has mainly been studied as a functional component of cornea, cartilage, and bone tissues. Changes to KS expression in the cornea have profound effects in diseases such as macular corneal dystrophy and keratoconus.¹⁰ Moreover, recent studies have also implicated KS in the inhibition of neural plasticity after spinal cord injury.⁹ KS polysaccharides are linear polymers of lactosamine, made up of galactose (Gal) and GlcNAc monosaccharides. Sulfation of KS has only been found to occur at the 6-*O* position of either Gal or GlcNAc, limiting its diversity as compared to heparin/HS and CS/DS. However, a unique feature of KS is its expression as both an *N*-linked (KS I) and *O*-linked glycan (KS II).

Biosynthesis of Glycosaminoglycans

The biosynthesis of GAGs involves the action of a variety of different Golgi-bound glycosyltransferases and sulfotransferases. Their activity leads to the appendage of an oligosaccharide linker onto the protein core, polymerization of disaccharide units, and specific sulfation of hydroxyl groups. By targeting these enzymes, studies have established the important roles of GAGs in different biological contexts and diseases. Therefore, modulation of these biosynthetic enzymes is of great interest in the pursuit of therapies for many illnesses.

Heparin, HS, CS, and DS biosynthesis initiates with the attachment of a tetrasaccharide linker region onto the core protein. This tetrasaccharide is always composed of GlcA-Gal-Gal-Xyl. Xylose (Xyl) is first appended to the core protein by the action of one of two xylosyltransferases, XylT1 or XylT2. Two Gal monosaccharides are then added sequentially by the galatosyltransferases GalT-1 and GalT-2, respectively. Lastly, completion of the linker is performed by the addition of GlcA by the GlcAT-1 enzyme. Besides genetic methods, this linker region can be taken advantage of to study GAGs by using β -xylosides as a way to prime GAG synthesis and compete with the endogenous protein bound Xyl.²²⁻²⁴

At this point, heparin/HS and CS/DS biosynthesis diverge by the addition of the GlcNAc or GalNAc, respectively, onto the tetrasaccharide linker. In heparin/HS biosynthesis, the family of exostosin-like glycosyltransferases (ExtL1, ExtL2, ExtL3) is responsible for the initiation of polymerization by the attachment of the GlcNAc residue. After this initiation, a heterodimer of the exostosin glycosyltransferases Ext1 and Ext2

polymerizes the rest of the polysaccharide. Deacetylation and sulfation can then begin on the growing polymer through the catalysis of *N*-deacetylase/*N*-sulfotransferase (Ndst). This is followed by epimerization of GlcA to IdoA by C5-epimerase. Further sulfation of the 2-*O*, 3-*O*, and 6-*O* positions of the disaccharide can then occur through the activity of the heparan 2-*O*-, 3-*O*-, and 6-*O*-sulfotransferases (Hs2st, Hs3st, and Hs6st). These deacetylation, sulfation, and epimerization steps of GAG chains are not found to be performed to completion, thus leading to heterogeneity in the polymer.²⁵⁻²⁷

CS/DS synthesis is triggered by the addition of GalNAc by *N*-acetylgalactosaminyltransferase-I (GalNAcT-I) to the core tetrasaccharide (Figure 1.3). Further elongation occurs with chondroitin synthetase (ChSy), a protein complex containing GalNAcTs, glucuronosyltransferases (GlcATs), and chondroitin polymerizing factor (ChPF). It has been shown that impairment of these chain polymerizing enzymes leads to the skeletal abnormalities such as chondrodysplasia, decreased bone density, and digit patterning defects. At this point, DS-epimerase can convert GlcA monosaccharides to IdoA and generate the DS disaccharide. However, like heparin and HS biosynthesis, this step is often substoichiometric leading to a linear polysaccharide with a mixture of GlcA and IdoA. On the polymerizing chain, sulfotransferases can then begin to differentially sulfate the 2-*O* position of GlcA or 4-*O* and 6-*O* positions of GalNAc. In CS/DS synthesis, there are distinct enzymes involved in the generation of each sulfation motif. As such, the synthesis of the CS/DS-A and CS/DS-C units by chondroitin/dermatan 4-*O*-sulfotransferases (Chst11-14) and chondroitin 6-*O*-sulfotransferase (Chst3), respectively, are prerequisites for the synthesis of the disulfated CS/DS-D and CS/DS-E units. The sulfotransferases responsible for this

activity are uronyl 2-*O*-sulfotransferase (Ust) and GalNAc 4S-6-*O*-sulfotransferase (Chst15) for the biosynthesis of CS/DS-D and CS/DS-E, respectively.^{11,28}

KS biosynthesis, unlike the other family of GAGs, begins with the appendage of an oligosaccharide core that depends on if it is *N*-linked (KS I) or *O*-linked (KS II). KS I shares the same core oligosaccharide as all human *N*-linked glycans, a Man-GlcNAc-GlcNAc trisaccharide that is core fucosylated on the reducing end GlcNAc. The nonreducing end mannose (Man) is further elaborated into a biantennary structure with two Man units. One branch is further elaborated into KS, while the other becomes a shorter oligosaccharide terminated with sialic acid. The KS II core oligosaccharide contains a singular GalNAc residue appended to serine or threonine sidechains that branches into GlcNAc and Gal. The KS linear polysaccharide is attached to the GlcNAc branch, while the Gal-containing antenna is usually terminated immediately by sialic acid. KS disaccharide units are polymerized by a collection of *N*-acetylglucosaminyltransferases (GlcNAcT-I, -II, and -VII) and galactosyltransferases (GalT-I-IV). Sulfation can then occur on the 6-*O* hydroxyl groups of both GlcNAc and Gal residues. GlcNAc sulfation is carried through by GlcNAc 6-*O*-sulfotransferases (Chst2, Chst4, Chst6), while Gal sulfation is catalyzed by the Gal 6-*O*-sulfotransferase (Chst1).^{10,29}

Glycosaminoglycan Sulfotransferases

As indicated before, the sulfation patterns of GAGs are tightly regulated by a large family of sulfotransferase enzymes. Two major classes of sulfotransferases, cytosolic and

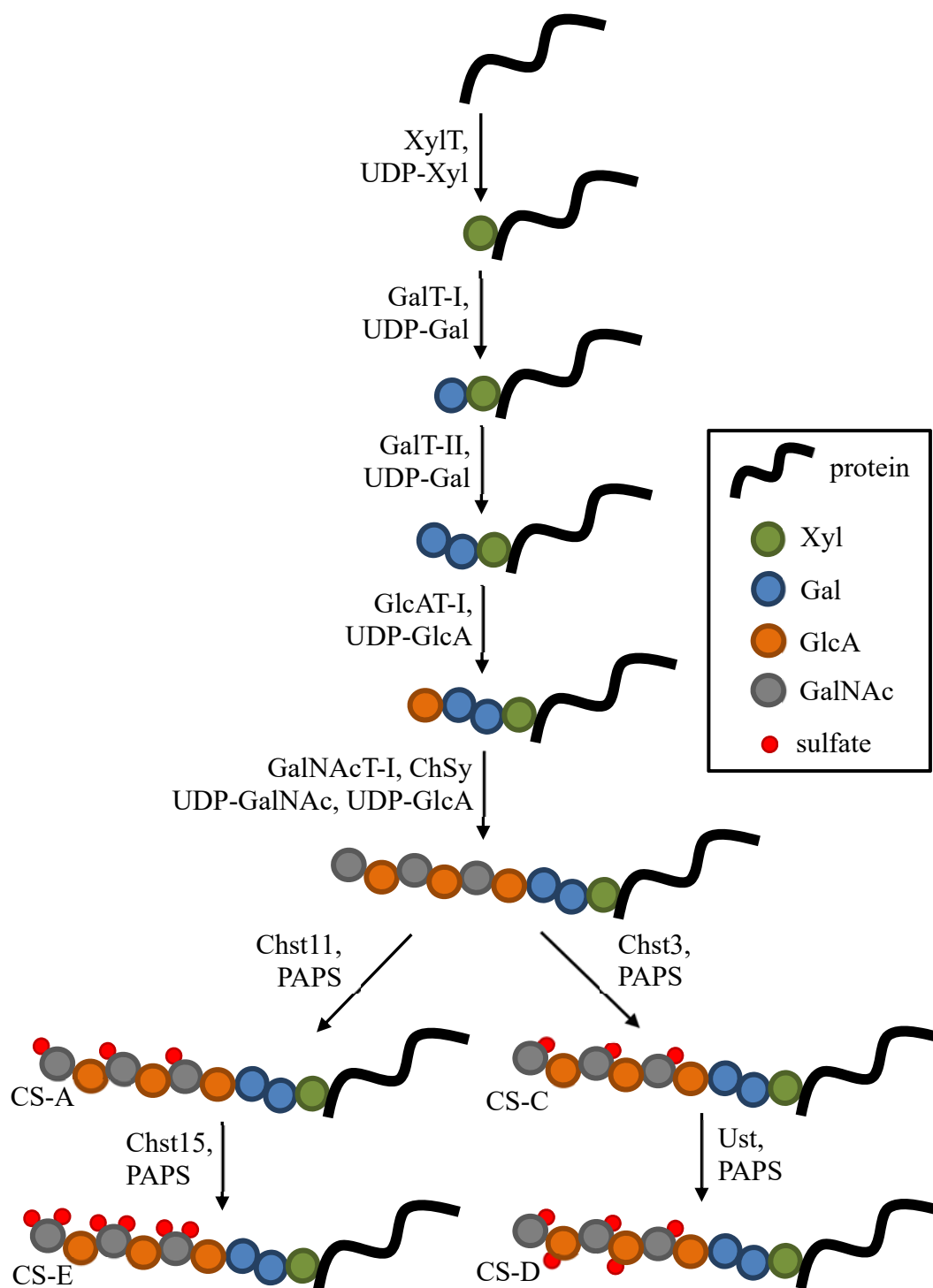


Figure 1.3. Illustrative representation of CS biosynthesis. Xyl, xylose; XylT, xylosyltransferase; Gal, galactose; GalT, galactosyltransferase; GlcA, glucuronic acid; GlcAT, glucuronyltransferase; GalNAc, *N*-acetylgalactosamine; GalNAcT, *N*-acetylgalactosaminyltransferase; ChSy, chondroitin synthetase; Chst11, CS 4-*O*-sulfotransferase; Chst3, CS 6-*O*-sulfotransferase; Chst15, GalNAc 4S-6-*O*-sulfotransferase; Ust, uronylsulfotransferase; PAPS, 3'-phosphoadenosine-5'-phosphosulfate.

membrane-associated, catalyze the transfer of a sulfonyl group from the donor molecule 3'-phosphoadenosine-5'-phosphosulfate (PAPS) to a variety of amine- and hydroxyl-containing substrates. Cytosolic sulfotransferases sulfonate small molecules, such as hormones, bioamines, drugs, and other xenobiotic agents. Membrane-associated sulfotransferases sulfonate larger biomolecules, including proteins and GAGs.³⁰

Specifically, GAG sulfotransferases are Golgi-membrane resident enzymes capable of sulfonating free hydroxyl or amine groups on the polymerizing GAG polysaccharide chain. Their overall structure comprises of a single pass transmembrane domain (with the exception of Hs3st), stem region, and luminal catalytic domain.³¹ The majority of known sulfotransferase crystal structures are of the cytosolic class. The limited information about GAG sulfotransferase structure has contributed to the difficulty of developing tools to study them. Genetic manipulation of these sulfotransferases have implicated them in many cellular functions, including cell migration,³² cancer metastasis,³³ and development.^{34,35} Thus, therapies and tools to interrogate GAG sulfation have the potential to be extremely useful.

Even though there are large differences between the acceptor substrates of the sulfotransferase family, they all have a sequence and structurally conserved PAPS-binding site. The structural features important for PAPS binding comprise of a 5'-phosphosulfate-binding (PSB) loop, 3'-phosphate-binding (PB) loop, and a β -strand-loop- α -helix (SLH3) motif that interacts with the adenosine ring.³⁶ Key amino acids often include hydrogen bond promoting and basic residues to counter balance negative charge. Due to the conserved nature of this domain and the similarity between PAPS and adenosine triphosphate (ATP), the phosphate donor of kinases, kinase inhibitor library scaffolds have been used as starting points for drug discovery against these enzymes.^{37,38}

Based on the structural similarities and utilization of the sulfuryl donor, PAPS, it is hypothesized that sulfotransferases share a common catalytic mechanism.³⁹ The mechanism that has been proposed for cytosolic sulfotransferases proceeds through an S_N2-like reaction whereby the acceptor hydroxyl acts as a nucleophile and attacks the sulfur atom of PAPS. This mechanism hinges on conserved basic residues to serve as a catalytic base and stabilize negative charge buildup on the leaving 3'-phosphoadenosine-5'-phosphate (PAP) group.³⁰ These residues are typically conserved histidine and lysine residues in cytosolic sulfotransferases. However, in Ndst and Hs3st, the catalytic histidine has been suggested to be replaced by a glutamate.⁴⁰ In order to rationally design chemical tools to study this family of enzymes, we will need to further understand their structure and mechanism.

Biological Role of Sulfation on Glycosaminoglycans

The importance of GAG sulfation has been known for decades. However, the complexities of the molecule have limited the ways it can be studied to genetic manipulations and challenging chemical syntheses. Nevertheless, some studies that have been reported have shown striking biological implications, including anti-coagulation, viral invasion, macular dystrophy, and spinal cord injury.

The sulfation of heparin and HS are critical in the biological contexts of anti-coagulation and viral invasion. The anti-coagulant drug activity of heparin is the result of the interaction between a specifically sulfated pentasaccharide sequence and antithrombin III. Removal of specific sulfate groups within this pentasaccharide abolishes its anti-coagulant activity.⁴¹ In another biological context, expression cloning experiments in cells resistant to

herpes simplex virus (HSV) invasion were used to show how HS 3-*O* sulfation was necessary for viral invasion.¹

Macular corneal dystrophy (MCD) is an autosomal recessive hereditary disease that results in a loss of vision and eventually necessitates a corneal transplant. It was discovered that patients with MCD have mutations and large genomic deletions/replacements in or before the encoding of Chst6, the sulfotransferase responsible for 6-*O* sulfation of GlcNAc residues in keratan sulfate. These changes were determined to be causative for the disease.⁴²

A transgenic mouse overexpressing the human 6-*O*-sulfotransferase (Chst3) revealed the effects of sulfation on cortical plasticity. The ratio of the two most abundant sulfation motifs, CS-A and CS-C, play a large role in determining the end of a critical period in ocular development. Transgenic mice with an increase in 6-*O* sulfation retained ocular dominance plasticity at a time that usually coincides with the end of the critical period of cortical plasticity. The decrease in sulfation prevented the accumulation of CS around neuronal synapses and localization of Otx2 homeoprotein, a regulator of interneuronal maturation.²¹

Our lab's long-standing interest in CS and its roles in neuroregeneration were inspired by its known inhibitory role in spinal cord injury. Studies have shown that digestion of CS at the site of injury with chondroitinase ABC (ChABC) leads to axon regeneration and functional recovery.^{8,43-45} We discovered that a specific sulfation motif, CS-E, was primarily responsible for the inhibitory effects of CS, and we determined that a monoclonal antibody against CS-E was capable of promoting optic nerve regeneration *in vivo*.⁴⁶ These studies suggested that blocking the CS-E motif might be an exciting, new strategy for neuronal repair.

Tools for Modulating Glycosaminoglycan Sulfation

While there are many tools to study the general role of GAGs in biological contexts (samples purified from natural sources, genetic models, antibodies, enzymes, small molecules), the methods to specifically modulate and study sulfation is far more limited. Cellular sulfation levels are typically manipulated genetically or by using sodium chlorate.

Genetically altering the expression of the sulfotransferases that regulate sulfation is a common technique to study GAG sulfation. It can allow for the manipulation of a particular sulfation motif, such as the example of the CS 6-*O* sulfotransferase discussed above. However, there are drawbacks of genetic knockdown and knockout models. Unknown effects of compensatory mechanisms can play a significant role in displayed phenotypes. There are many sulfotransferases with redundant activity in the GAG biosynthetic pathway which can play a role in limiting the usefulness of information from these models. Additionally, from a therapeutic perspective, methods to manipulate gene expression are still not yet viable.

Alternatively, small molecules benefit from their ease of use and potential for therapeutic development. Sodium chlorate is a simple molecule that inhibits the adenosine triphosphate (ATP) sulfurylase activity of PAPS synthetase, the enzyme that produces PAPS. Crystallographic data has previously suggested that inhibition occurs in a noncompetitive fashion by the blocking of the release of the pyrophosphate and adenosine 5'-phosphosulfate (APS) products from the enzyme active site.⁴⁷ Inhibition of ATP sulfurylase activity globally affects all types of sulfation and thus has a multitude of potential off-target effects beyond GAG sulfation. It is typically used at high *in vitro* working concentrations in the range of 10

to 50 mM. Furthermore, along with its millimolar effective concentration, its oxidative potential and toxicity preclude it from being used therapeutically.^{48,49} The ability to selectively target GAG sulfotransferases with small molecules would provide a valuable set of new chemical genetic tools for manipulating and understanding GAGs.

REFERENCES

- (1) Shukla, D.; Liu, J.; Blaiklock, P.; Shworak, N. W.; Bai, X.; Esko, J. D.; Cohen, G. H.; Eisenberg, R. J.; Rosenberg, R. D.; Spear, P. G. *Cell* **1999**, *99*, 13.
- (2) Chen, Y.; Maguire, T.; Hileman, R. E.; Fromm, J. R.; Esko, J. D.; Linhardt, R. J.; Marks, R. M. *Nat. Med.* **1997**, *3*, 866.
- (3) Sasisekharan, R.; Shriver, Z.; Venkataraman, G.; Narayanasami, U. *Nat. Rev. Cancer* **2002**, *2*, 521.
- (4) Vallen, M. J.; Schmidt, S.; Oosterhof, A.; Bulten, J.; Massuger, L. F.; van Kuppevelt, T. H. *Plos One* **2014**, *9*, e111806.
- (5) Li, F.; ten Dam, G. B.; Murugan, S.; Yamada, S.; Hashiguchi, T.; Mizumoto, S.; Oguri, K.; Okayama, M.; van Kuppevelt, T. H.; Sugahara, K. *J. Biol. Chem.* **2008**, *283*, 34294.
- (6) Maeda, N. *Front. Neurosci.* **2015**, *9*, 98.
- (7) Carulli, D.; Laabs, T.; Geller, H. M.; Fawcett, J. W. *Curr. Opin. Neurobiol.* **2005**, *15*, 116.
- (8) Bradbury, E. J.; Moon, L. D. F.; Popat, R. J.; King, V. R.; Bennett, G. S.; Patel, P. N.; Fawcett, J. W.; McMahon, S. B. *Nature* **2002**, *416*, 636.
- (9) Imagama, S.; Sakamoto, K.; Tauchi, R.; Shinjo, R.; Ohgomori, T.; Ito, Z.; Zhang, H.; Nishida, Y.; Asami, N.; Takeshita, S.; Sugiura, N.; Watanabe, H.; Yamashita, T.; Ishiguro, N.; Matsuyama, Y.; Kadomatsu, K. *J. Neurosci.* **2011**, *31*, 17091.
- (10) Funderburgh, J. L. *Glycobiology* **2000**, *10*, 951.
- (11) Sugahara, K.; Mikami, T.; Uyama, T.; Mizuguchi, S.; Nomura, K.; Kitagawa, H. *Curr. Opin. Struc. Biol.* **2003**, *13*, 612.

- (12) Shriver, Z.; Capila, I.; Venkataraman, G.; Sasisekharan, R. *Handb. Exp. Pharmacol.* **2012**, 159.
- (13) Petitou, M.; Herault, L. P.; Bernat, A.; Driguez, P. A.; Duchaussoy, P.; Lormeau, J. C.; Herbert, J. M. *Nature* **1999**, 398, 417.
- (14) Perrimon, N.; Bernfield, M. *Nature* **2000**, 404, 725.
- (15) Sasisekharan, R.; Ernst, S.; Venkataraman, G. *Angiogenesis* **1997**, 1, 45.
- (16) Gallagher, J. T.; Walker, A. *Biochem. J.* **1985**, 230, 665.
- (17) Kim, S. H.; Turnbull, J.; Guimond, S. *J. Endocrinol.* **2011**, 209, 139.
- (18) Hwang, H. Y.; Olson, S. K.; Esko, J. D.; Horvitz, H. R. *Nature* **2003**, 423, 439.
- (19) Mizuguchi, S.; Uyama, T.; Kitagawa, H.; Nomura, K. H.; Dejima, K.; Gengyo-Ando, K.; Mitani, S.; Sugahara, K.; Nomura, K. *Nature* **2003**, 423, 443.
- (20) Gama, C. I.; Tully, S. E.; Sotogaku, N.; Clark, P. M.; Rawat, M.; Vaidehi, N.; Goddard, W. A., 3rd; Nishi, A.; Hsieh-Wilson, L. C. *Nat. Chem. Biol.* **2006**, 2, 467.
- (21) Miyata, S.; Komatsu, Y.; Yoshimura, Y.; Taya, C.; Kitagawa, H. *Nat. Neurosci.* **2012**, 15, 414.
- (22) Schwartz, N. B.; Galligani, L.; Ho, P. L.; Dorfman, A. *Proc. Natl. Acad. Sci. U. S. A.* **1974**, 71, 4047.
- (23) Lugemwa, F. N.; Esko, J. D. *J. Biol. Chem.* **1991**, 266, 6674.
- (24) Mizumoto, S.; Ikegawa, S.; Sugahara, K. *J. Biol. Chem.* **2013**, 288, 10953.
- (25) Kreuger, J.; Kjellen, L. *J. Histochem. Cytochem.* **2012**, 60, 898.
- (26) Multhaupt, H. A.; Couchman, J. R. *J. Histochem. Cytochem.* **2012**, 60, 908.
- (27) Sasisekharan, R.; Venkataraman, G. *Curr. Opin. Chem. Biol.* **2000**, 4, 626.
- (28) Mikami, T.; Kitagawa, H. *Biochim. Biophys. Acta* **2013**, 1830, 4719.
- (29) Pomin, V. H. *Int. J. Biol. Macromolec.* **2015**, 72, 282.
- (30) Chapman, E.; Best, M. D.; Hanson, S. R.; Wong, C. H. *Angew. Chem. Int. Edit.* **2004**, 43, 3526.
- (31) Kusche-Gullberg, M.; Kjellen, L. *Curr. Opin. Struc. Biol.* **2003**, 13, 605.

- (32) Nikolovska, K.; Spillmann, D.; Seidler, D. G. *J. Cell Sci.* **2015**, *128*, 460.
- (33) Mizumoto, S.; Watanabe, M.; Yamada, S.; Sugahara, K. *Biomed. Res. Int.* **2013**.
- (34) Bink, R. J.; Habuchi, H.; Lele, Z.; Dolk, E.; Joore, J.; Rauch, G. J.; Geisler, R.; Wilson, S. W.; den Hertog, J.; Kimata, K.; Zivkovic, D. *J. Biol. Chem.* **2003**, *278*, 31118.
- (35) Ringvall, M.; Kjellen, L. *Prog. Mol. Biol. Transl.* **2010**, *93*, 35.
- (36) Raman, R.; Myette, J.; Venkataraman, G.; Sasisekharan, V.; Sasisekharan, R. *Biochem. Bioph. Res. Co.* **2002**, *290*, 1214.
- (37) Armstrong, J. I.; Portley, A. R.; Chang, Y. T.; Nierengarten, D. M.; Cook, B. N.; Bowman, K. G.; Bishop, A.; Gray, N. S.; Shokat, K. M.; Schultz, P. G.; Bertozzi, C. R. *Angew. Chem. Int. Edit.* **2000**, *39*, 1303.
- (38) Chapman, E.; Ding, S.; Schultz, P. G.; Wong, C. H. *J. Am. Chem. Soc.* **2002**, *124*, 14524.
- (39) Gorokhov, A.; Perera, L.; Darden, T. A.; Negishi, M.; Pedersen, L. C.; Pedersen, L. G. *Biophys. J.* **2000**, *79*, 2909.
- (40) Edavettal, S. C.; Lee, K. A.; Negishi, M.; Linhardt, R. J.; Liu, J.; Pedersen, L. C. *J. Biol. Chem.* **2004**, *279*, 25789.
- (41) Vanboeckel, C. A. A.; Petitou, M. *Angew. Chem. Int. Edit.* **1993**, *32*, 1671.
- (42) Akama, T. O.; Nishida, K.; Nakayama, J.; Watanabe, H.; Ozaki, K.; Nakamura, T.; Dota, A.; Kawasaki, S.; Inoue, Y.; Maeda, N.; Yamamoto, S.; Fujiwara, T.; Thonar, E. J.; Shimomura, Y.; Kinoshita, S.; Tanigami, A.; Fukuda, M. N. *Nat. Genet.* **2000**, *26*, 237.
- (43) Garcia-Alias, G.; Barkhuysen, S.; Buckle, M.; Fawcett, J. W. *Nat. Neurosci.* **2009**, *12*, 1145.
- (44) Wang, D.; Ichiyama, R. M.; Zhao, R.; Andrews, M. R.; Fawcett, J. W. *J. Neurosci.* **2011**, *31*, 9332.
- (45) Alilain, W. J.; Horn, K. P.; Hu, H.; Dick, T. E.; Silver, J. *Nature* **2011**, *475*, 196.
- (46) Brown, J. M.; Xia, J.; Zhuang, B. Q.; Cho, K. S.; Rogers, C. J.; Gama, C. I.; Rawat, M.; Tully, S. E.; Uetani, N.; Mason, D. E.; Tremblay, M. L.; Peters, E. C.; Habuchi, O.; Chen, D. F.; Hsieh-Wilson, L. C. *P. Natl. Acad. Sci. U.S.A.* **2012**, *109*, 4768.
- (47) Ullrich, T. C.; Huber, R. *J. Mol. Biol.* **2001**, *313*, 1117.

(48) Baeuerle, P. A.; Huttner, W. B. *Biochem. Biophys. Res. Co.* **1986**, *141*, 870.

(49) Steffen, C.; Wetzel, E. *Toxicology* **1993**, *84*, 217.

Chapter 2

CHST15-ANTAGONIST HIGH-THROUGHOUT SCREEN AND VALIDATION

Chapter 2

CHST15-ANTAGONIST HIGH-THROUGHPUT SCREEN AND VALIDATION

Introduction

The ability to selectively target GAG sulfotransferases with small molecules would provide a valuable set of new chemical genetic tools for manipulating and understanding GAGs. Such molecules will also be essential for controlling GAG-mediated processes involved in various diseases. Notably, GAG sulfation patterns play an important role in the progression and treatment of diseases such as spinal cord injury,¹⁻³ Alzheimer's disease,⁴ osteoarthritis,^{5,6} cancer,^{7,8} malaria,^{9,10} herpes,¹¹ and macular dystrophy.^{12,13} One approach to discover novel scaffolds that can inhibit these enzymes is through high-throughput screening. Here, we developed and optimized a fluorescent high-throughput screen to identify small molecule inhibitors against Chst15. The assay was miniaturized and screened against a 70,000 compound library. Hit compounds were validated by secondary radioisotope labeling assays for further lead development.

Results and Discussion

To begin to identify small molecules that could inhibit Chst15 activity, we opted to develop a fluorescent, enzyme coupled screen. This screen was based around a previously reported assay utilizing a PAPS regeneration cycle (Figure 2.1).^{14,15} In this assay, the coupled enzyme, aryl sulfotransferase IV (Sult1c1), has the ability to convert PAP to PAPS

by using 4-methylumbelliferyl sulfate (MUS) as the sulfate donor. Fluorescent 4-methylumbelliferone (MU) is generated by the liberation of this sulfate moiety from MUS. In this way, since Chst15 utilizes PAPS to generate CS-E, disruption of this PAPS regeneration cycle would allow us to effectively monitor inhibition levels.

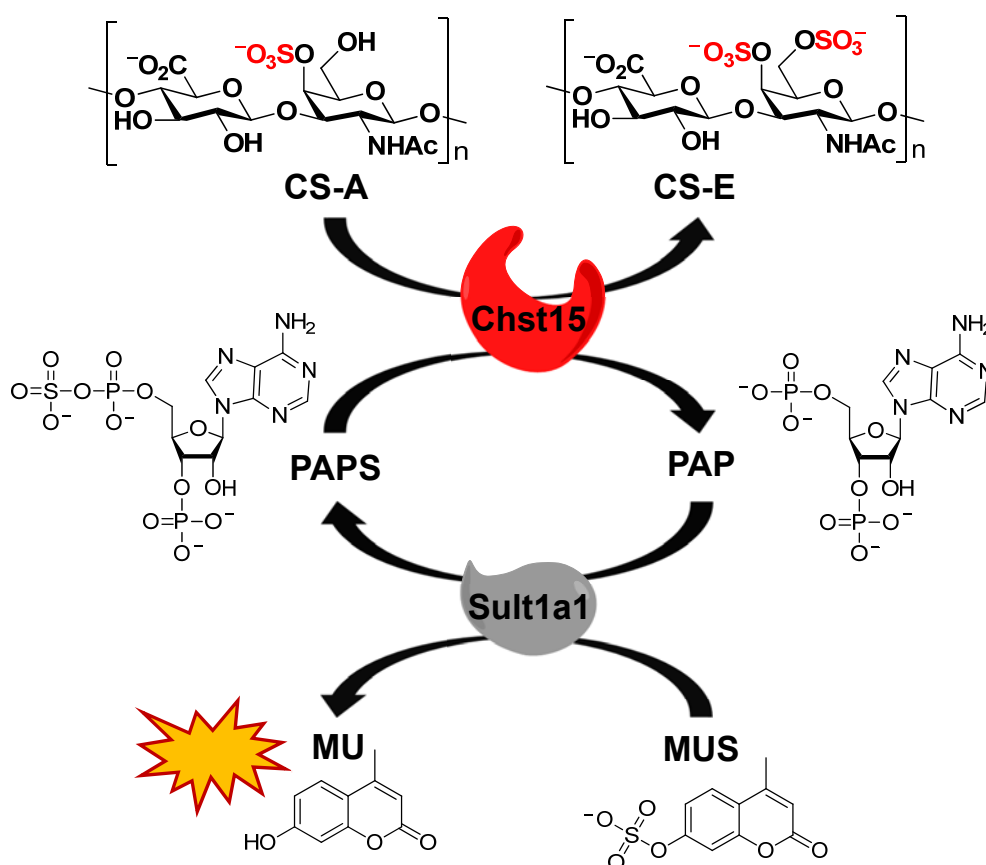


Figure 2.1. High-throughput assay for the discovery of Chst15 small molecule inhibitors. Inhibition of Chst15 was monitored through a fluorescent, enzyme-coupled readout. Fluorescence was generated by the conversion of MUS to MU by Sult1c1, which is linked to Chst15 via a PAPS regeneration cycle.

To develop such an assay, we began by bacterially expressing a maltose binding protein (MBP) fusion of the luminal catalytic domain of Chst15. After affinity purification, we recognized that the protein was co-purified with the chaperone proteins and possible truncations. We ensured that sulfotransferase activity was the result of the intended construct by designing and expressing K263A, R393A, and S400A mutants of Chst15. These designed mutations were all within the conserved PAPS binding domain. As expected, the mutations greatly attenuated activity when tested by a radioisotope labeling assay, indicating that the sulfotransferase activity witnessed from the wild type (WT) construct was the source of sulfotransferase activity (Figure 2.2).

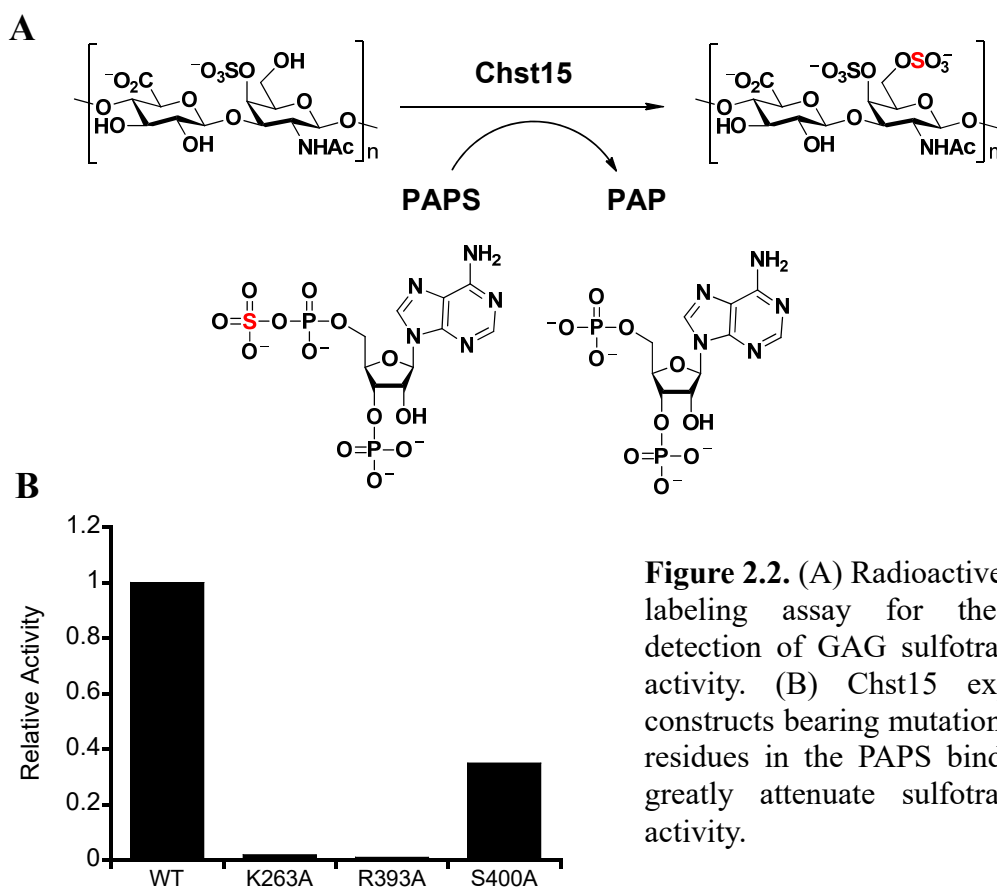


Figure 2.2. (A) Radioactive isotope labeling assay for the direct detection of GAG sulfotransferase activity. (B) Chst15 expression constructs bearing mutations to key residues in the PAPS binding site greatly attenuate sulfotransferase activity.

With an active enzyme in hand, we sought to test and optimize the fluorescent assay in a 96-well plate format. Two parameters that are critical for a high quality assay are a large signal-to-noise ratio (SNR) and a Z-factor (Z') greater than 0.5. We posited that the factors that would be most important for maximizing these parameters were pre-incubation time and enzyme concentrations. When Sult1c1 is bacterially expressed, it is known to co-purify with PAP tightly bound in the active site.^{16,17} Any PAP impurities at the start of the assay reaction would lead to immediate generation of fluorescent signal that is not a result of Chst15 activity. Therefore, it was essential to pre-incubate Sult1c1 with the other components of the assay minus Chst15 and allow full conversion of PAP impurities to PAPS. This starting background signal can then be subtracted from the overall increase after the addition of Chst15. We determined that after a 1 h pre-incubation time, the increase in background signal as a result of PAP impurities had begun to plateau and would negligibly affect the signal to noise ratio (Figure 2.3A).

The next assay components that were optimized were Sult1c1 and Chst15 enzyme concentrations. A high concentration of Sult1c1 was necessary for the immediate conversion of PAP to PAPS. This way, the rate limiting factor in the assay would be Chst15 activity and thus, maximize the SNR ratio. In the end, we were able to increase Sult1c1 concentrations to 0.1 $\mu\text{g}/\mu\text{L}$, which limited by protein stability and final glycerol concentrations (Figure 2.3B). Next, when Chst15 concentration was varied, we noticed a strong increase in fluorescent signal with a higher concentrations, similar to the effects of Sult1c1. From this data, we believed that the SNR and Z' could be tuned after miniaturization by varying the enzyme concentration (Figure 2.3C). The SNR and Z' at a

Chst15 concentration of $0.5 \mu\text{g}/\mu\text{L}$ was 6.14 and 0.62, respectively, after 15 min, well within the optimal assay range.

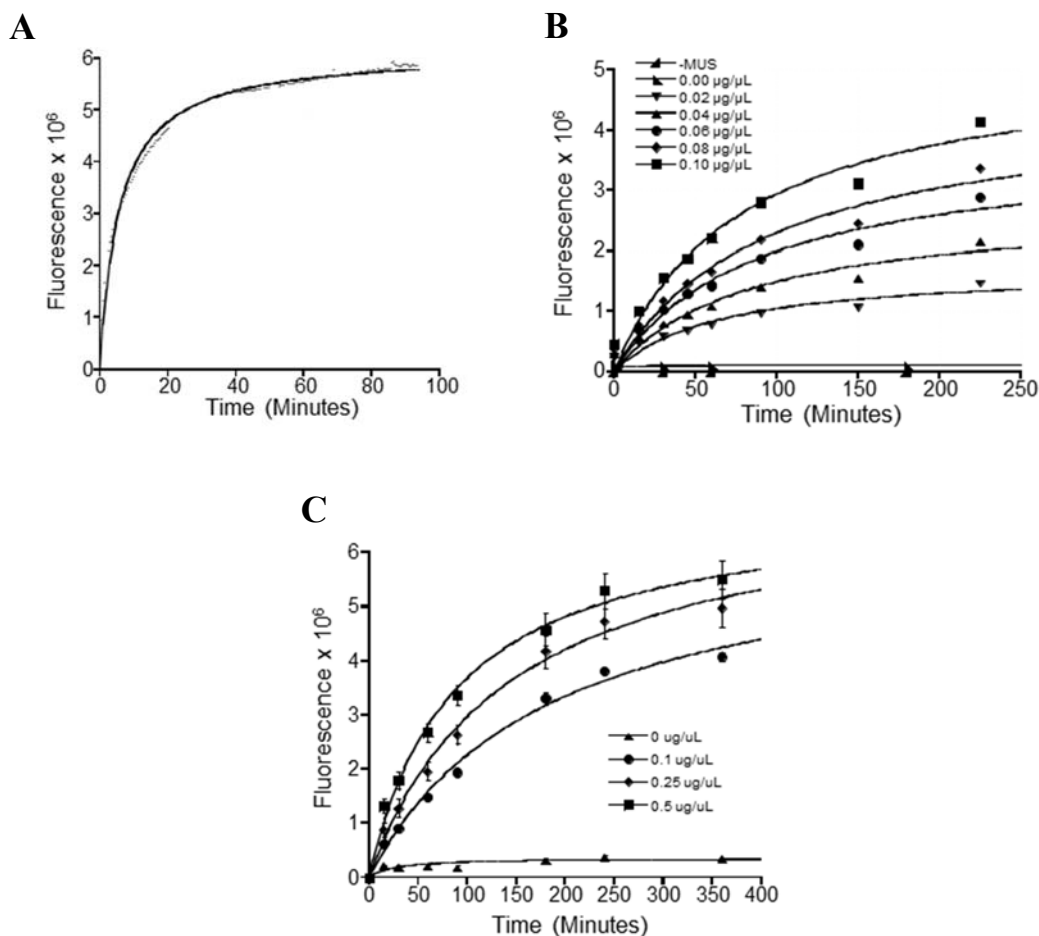


Figure 2.3. (A) Fluorimetry revealed an optimal pre-incubation time of 1 h to remove background fluorescence caused by PAP impurities. (B) Increasing Sult1c1 concentration, to a maximal of $0.1 \mu\text{g}/\mu\text{L}$, led to robust conversion of PAP to PAPS. (C) Increasing the concentration of Chst15 also led to greater fluorescence. Larger amounts of Chst15 could be used to increase the assay SNR and Z'.

Another important factor that plays key role in the quality of hits gained from the screen includes the concentration of the enzyme substrates. Assaying at the K_m of both

substrates enables the best coverage of both competitive and uncompetitive inhibitors.¹⁸ We used fluorimetry to monitor initial reaction rates at varying substrate concentrations under pseudo-first order conditions. The K_m 's for CS-A and PAPS were determined to be 1.11 mM and 32.91 μ M, respectively (Figure 2.4).

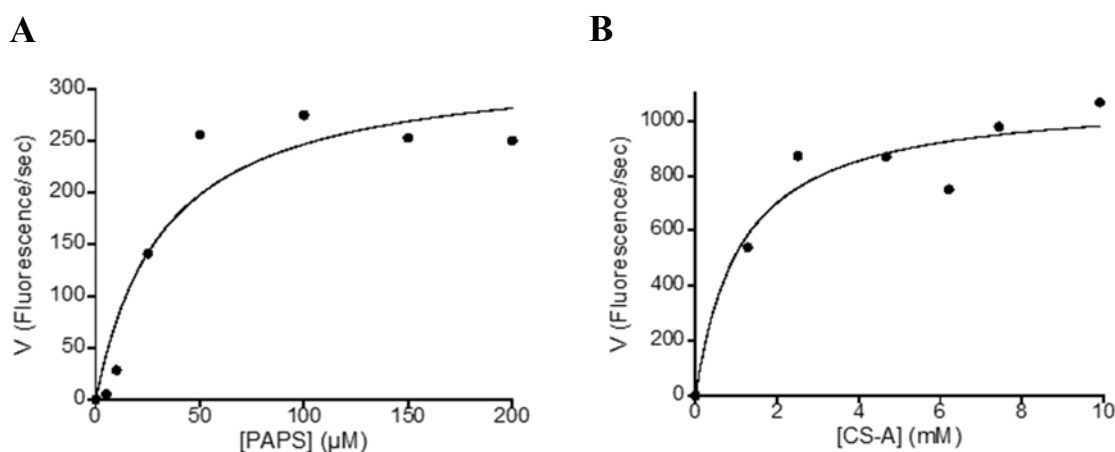


Figure 2.4. The K_m of PAPS and CS-A were determined to be 32.91 μ M and 1.11 mM, respectively, for Chst15 under optimized assay conditions.

With an optimized assay in hand, in collaboration with the Genomics Institute of the Novartis Research Foundation (GNF), we were able to miniaturize the assay to a 1536-well plate format. For the high-throughput assay, Chst15 was expressed in HEK293F cells as it was more feasible to generate enough active enzyme necessary to run the screen. The Z' for the miniaturized assay was 0.5, sufficient for screening. A diversity set of 70,000 small molecules were screened at 12.5 μ M to determine if they could decrease the detectable levels of fluorescence. We additionally counterscreened any small molecule that decreased the fluorescence of the complete, enzyme-coupled assay against Sult1c1 alone

to eliminate nonspecific inhibitors. After both initial screens, 26 compounds emerged to inhibit Chst15 with low Sult1c1 inhibition.

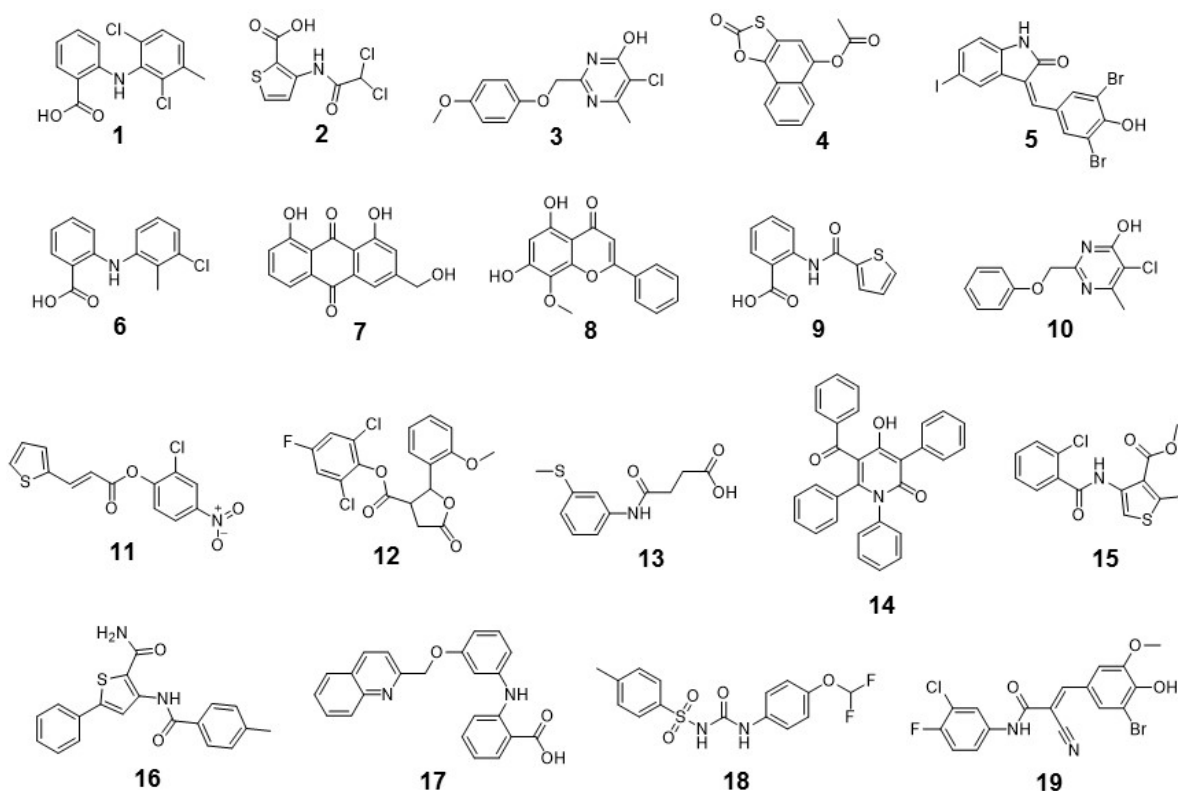


Figure 2.5. Structures of 19 of the 26 small-molecule hits that were commercially available for further evaluation.

Further validation of the hits were performed by reassessing in a 96-well plate format as well as the same direct, sensitive radioactive assay used previously (Figure 2.2A). First, 19 of the 26 hits (Figure 2.5) were obtained commercially and retested in the counterscreen in a 96-well plate format. It was determined that 12 of the 22 hits actually inhibited Sult1c1 (Figure 2.6A). The 7 hits that did not were carried through to the full

fluorescence assay. 5 of the 7 hits showed appreciable inhibitory activity specific for Chst15 at 12.5 μ M (Figure 2.6B).

However, drawbacks to the fluorescence assay include the complications that may arise from a system of indirect measurement of Chst15 activity and that the method eliminates potential lead scaffolds that may inhibit both Chst15 and Sult1c1. Therefore, to eliminate these complications, we opted to also test the hit compounds in the radioactive isotope labeling assay. 2 of the 19 hits showed inhibitory activity at high concentrations of 100 μ M when analyzed in this format (Figure 2.7A). The inhibitory activity of these small molecules, **5** and **19**, were further confirmed through dose response experiments. The IC_{50} of compounds **5** and **19** were determined to be 23 μ M and 39 μ M, respectively (Figure 2.7B). We surmised that this assay was more indicative of a working Chst15 inhibitor because it was a direct measurement of activity. Therefore, these two lead scaffolds were moved on for further development.

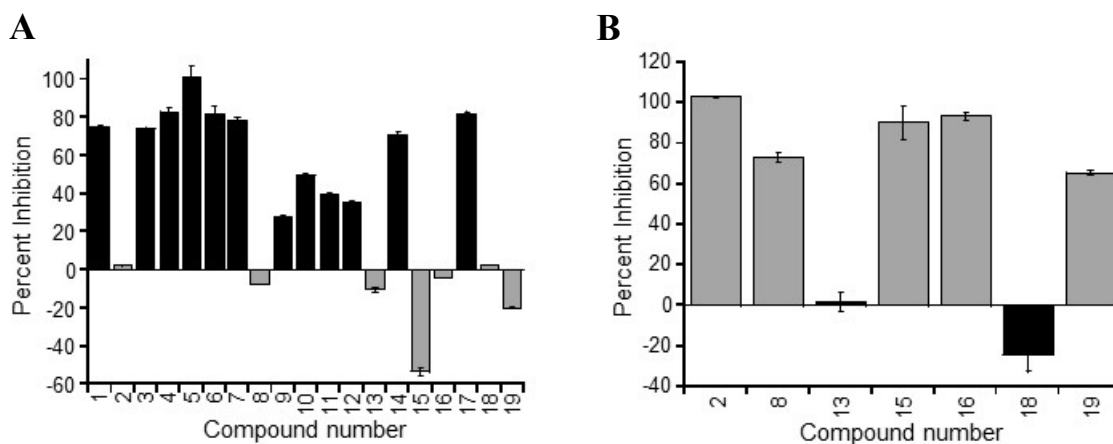


Figure 2.6. 19 of 26 available compounds were obtained and retested for Sult1c1 inhibition via the fluorescence assay. (A) 7 of 19 compounds did not appreciably inhibit the Sult1c1 at 12.5 μ M. (B) The 7 compounds that did not inhibit Sult1c1 were retested in the full fluorescence assay. 5 of 7 compounds inhibited Chst15 at 12.5 μ M. Values shown are the mean \pm S.E.M. ($n = 3$).

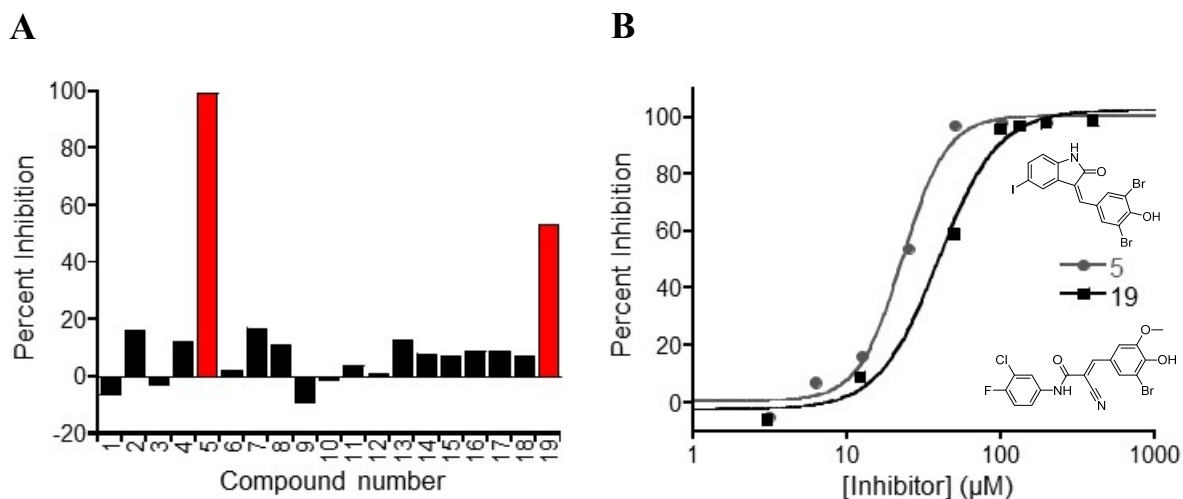


Figure 2.7. Validation of candidate compounds identified in the ~70,000 compound primary screen leads to two scaffolds. (A) 19 of 26 lead compounds were obtained and tested directly against Chst15 in a radioisotope labeling assay at 100 μM to confirm the inhibitory activity of two compounds, **5** and **19**. (B) Dose dependent inhibition was seen with compounds **5** ($\text{IC}_{50} = 23 \mu\text{M}$) and **19** ($\text{IC}_{50} = 39 \mu\text{M}$). All values reported as the mean of duplicate experiments.

Conclusion

In conclusion, here we designed and optimized a fluorescent, enzyme-coupled high-throughput screen for small molecule inhibitors against Chst15. Using this assay, we screened 70,000 molecules in collaboration with GNF. Of the 26 putative hits, 2 hits, compounds **5** and **10**, were confirmed to have dose dependent, micromolar potency against the enzyme *in vitro*. It will be important to further optimize and characterize these scaffolds to generate inhibitors with an inhibitory activity, specificity, and pharmacokinetic profile favorable for *in vivo* studies. This fluorescent screen should also be amenable for the evaluation of all cytosolic and membrane-bound sulfotransferases. Therefore, it could be designed for the discovery of small-molecule modulators of other GAG sulfotransferases to potentially attain a suite of inhibitors specific for different sulfation motifs.

Experimental Methods

Expression and Purification of Chst15. Origami B(DE3) (Novagen) harboring the pGro7 (TaKaRa) and pMAL-c2X-mChst15₁₅₁₋₅₆₁ constructs were kindly provided by Dr. Jian Liu (UNC-Chapel Hill). Cells were grown in LB medium (1L) at 37°C. When cells reached an OD₆₀₀ of 0.8, cells were moved to 22°C for 30 minutes. Isopropyl β -D-1-thiogalactopyranoside (IPTG) and L-(+)-arabinose was added to the cells at the final concentrations of 0.2 mM and 1 mg/mL, respectively, after the 30 minutes and cells were incubated for an additional 18 hours at 22°C. The pelleted cells were lysed in a cold buffer containing 20 mM Tris pH 7.5, 500 mM NaCl, 1 mM EDTA, and 1x CompleteTM protease inhibitors (Roche) by sonication. After centrifugation, the clarified lysate was added to pre-equilibrated amylose resin beads (New England Biolabs) and incubated at 4°C for 1 hour with end-over-end rotation. The beads were then washed 4 times with a buffer containing 20 mM Tris pH 7.5, 500 mM NaCl, and 1 mM EDTA and eluted with a buffer containing 20 mM Tris pH 7.5, 500 mM NaCl, 1 mM EDTA, and 10 mM D-(+)-maltose. After SDS-PAGE analysis, the purified protein was concentrated with a 50 kDa molecular weight cut-off (MWCO) spin filter (Millipore), buffer exchanged into 20 mM Tris pH 7.5, 500 mM NaCl, and 1 mM EDTA, and 50% glycerol, and stored at -20°C.

Expression and Purification of Sult1c1. BL21(DE3)RIL (Stratagene) harboring the pET28a-Sult1c1 was kindly provided by Dr. Jian Liu (UNC-Chapel Hill). Cells were grown in LB medium (1L) at 37°C. When cells reached an OD₆₀₀ of 0.8, cells were moved to 22°C for 30 minutes. Isopropyl β -D-1-thiogalactopyranoside (IPTG) was added to the

cells at the final concentrations of 0.2 mM after the 30 minutes and cells were incubated for an additional 18 hours at 22°C. The pelleted cells were lysed in a cold buffer containing 25 mM Tris pH 7.5, 500 mM NaCl, 10 mM imidazole, and 1x EDTA-free Complete™ protease inhibitors (Roche) by sonication. After centrifugation, the clarified lysate was added to pre-equilibrated Ni-NTA resin beads (Qiagen) and incubated at 4°C for 1 hour with end-over-end rotation. The beads were then washed 3 times with a buffer containing 25 mM Tris pH 7.5, 500 mM NaCl, and 10 mM imidazole, washed 1 time with 25 mM Tris pH 7.5, 500 mM NaCl, and 50 mM imidazole and eluted with a buffer containing 25 mM Tris pH 7.5, 500 mM NaCl, and 250 mM imidazole. After SDS-PAGE analysis, the purified protein was concentrated with a 30,000 Da molecular weight cut-off (MWCO) spin filter (Millipore), buffer exchanged into 25 mM Tris pH 7.5, 500 mM NaCl, and 50% glycerol, and stored at -20°C.

Enzymatic Synthesis and Purification of PAPS and ³⁵S-PAPS. Adenosine 5'triphosphate (ATP, 11.8 mM), MgCl₂ (11.8 mM), phosphoenolpyruvic acid (PEP, 24 mM), and NaSO₄ (118 mM) were dissolved into 137 µL of 50 mM Tris pH 8.0. If ³⁵S-PAPS is to be synthesized, 25 µL of ³⁵S-NaSO₄ (10 mCi/mL stock, American Radiolabeled Chemicals Inc.) was added to the reaction. If not, 25 µL of water was added instead. After, 1 µL each of pyruvate kinase (700 units/mL, Sigma), inorganic pyrophosphatase (250 units/mL, Sigma), ATP sulfurylase (2 units/mL, Sigma), and APS kinase (5 mg/mL, generously donated from Dr. Andrew Fisher (UC-Davis)) were added to the reaction mixture. The reaction was allowed to incubate at 37°C for 9 hours. After 9 hours, an

additional 1 μ L of the same four enzymes were added and the reaction was allowed to proceed an additional 12 hours at 37°C. Purification of PAPS was achieved using DEAE-cellulose (Whatman) column (2.5 mL resin) pre-equilibrated with water and eluted using a gradient of 0-500 mM NaCl. Fractions were lyophilized, reconstituted in 170 μ L of 1 mM Tris pH 8.0, and analyzed with HPLC using a Zorbax-NH₂ column (4.6 mm x 150 mm, Agilent). The following method was applied: 100% water for 10 minutes, followed by a linear gradient of 0-100% 1 M KH₂PO₄ for 30 minutes, followed by 100% 1 M KH₂PO₄ for 15 minutes at a flow rate of 1 mL/min with UV 254 nm detection. Fractions containing PAPS were stored at -80°C.

96-Well Plate Enzyme-Coupled, Fluorescent Assay. For full assay, the substrates PAPS and CS-A (Sigma) were added to a solution containing 100 mM Tris pH 7.6, 1 mM β -mercaptoethanol, and 5 mM 4-methylumbelliferyl sulfate at a concentration of 30 μ M and 1 mM, respectively, to a total volume of 74 μ L. 5 μ L of Sult1c1 (2 mg/mL) was added to the mixture and incubated at 37°C for 1 hour. 20 μ L of Chst15 (1 mg/mL), pre-incubated with 1 μ L of DMSO or inhibitor stock for 30 minutes, was then added to the mixture to initiate the reaction. The 96-well plate was read using either a Victor3 Plate Reader (Perkin Elmer) or Flexstation 3 (Molecular Devices) at excitation and emission wavelengths of 360 nm and 449 nm respectively.

For the counter screen, the substrates PAPS and CS-A (Sigma) were also added to a solution containing 100 mM Tris pH 7.6, 1 mM β -mercaptoethanol, and 5 mM 4-methylumbelliferyl sulfate at a concentration of 30 μ M and 1 mM, respectively, to a total volume of 75 μ L. The reaction was initiated by adding a mixture of 5 μ L of Sult1c1, 19

μ L of 100 mM Tris pH 7.6, and 1 μ L of DMSO or inhibitor stock that has been pre-incubated at room temperature for 30 minutes.

1536-Well Plate High-Throughput Screening Protocol. For the full assay, 2 μ L of Chst15 (0.46 mg/mL) was added to each well in black 1526-well plates pre-plated with the small molecule library (12.5 μ M final concentration). The mixture was incubated for 5 minutes with shaking. Then, 2 μ L of a mixture containing 100 mM Tris pH 7.6, 1 mM β -mercaptoethanol, 5 mM 4-methylumbelliferyl sulfate, 1 mM CS-A, 30 μ M PAPS, and 0.1 mg/mL Sult1a1 was added to the plate and read as the 0 minute time point on a PheraStar Plate Reader with an excitation wavelength of 350 nm and emission wavelength of 460 nm. The plate was incubated at 37°C for 30 minutes and was read again at the same wavelengths. For the counter screen, Chst15 was omitted.

Chst15 Radioisotope Labeling Assay. Chst15 activity was determined by incubating Chst15 with 150 μ g of CS-A (Sigma) and 30 μ M 35 S-PAPS in 150 μ L of reaction buffer (100 mM Tris pH 7.6, 5 mM β -mercaptoethanol) for 30 minutes at 37°C. For assaying the potency of inhibitors, 1.25 μ L of a DMSO stock was pre-incubated with Chst15 alone for 1 hour at room temperature prior to the addition of the substrates in reaction buffer. The samples were then subjected to DEAE-Sepharose chromatography (200 μ L slurry per reaction) to purify the 35 S-labeled product. Samples were washed 3 times with 1 mL of a buffer containing 50 mM NaOAc pH 5.5, 150 mM NaCl, 6 M urea, 1 mM EDTA, and 0.01% Triton X-100, washed 2 times with 1 mL of a buffer containing 50 mM NaOAc pH

5.5, 250 mM NaCl, and 0.01% Triton X-100, and eluted with 1 mL of a buffer containing 50 mM NaOAc pH 5.5, 1 M NaCl, and 0.01% Triton X-100. The elution was mixed with 7 mL of Ecoscint XR scintillation fluid (National Diagnostics) and activity was quantified with a liquid scintillation counter (Beckman LS6500).

Assay Optimization with Fluorimetry. The substrates PAPS and CS-A (Sigma) were added to a solution containing 100 mM Tris pH 7.6, 1 mM β -mercaptoethanol, and 5 mM 4-methylumbelliferyl sulfate at a concentration of 30 μ M and 1 mM, respectively, to a total volume of 75 μ L. 5 μ L of Sult1a1 (2 mg/mL) was added to the mixture and incubated at 37°C for 1 hour. 20 μ L of Chst15 (1 mg/mL) was then added to the mixture to initiate the reaction. The reaction mixture was added to a sub-micro fluorimeter cell (Starna Cells) and product formation was monitored for 10 minutes at excitation and emission wavelengths of 360 nm and 449 nm respectively using a steady-state fluorimeter (Caltech BILRC). K_m of the substrates was determined by varying substrate PAPS (10 μ M to 200 μ M) or CS-A (0.5 mM to 10 mM) concentrations and nonlinear regression analysis of initial velocity vs substrate concentration using Kaleidagraph (version 4.1.2).

REFERENCES

- (1) Carulli, D.; Laabs, T.; Geller, H. M.; Fawcett, J. W. *Curr. Opin. Neurobiol.* **2005**, *15*, 116.
- (2) Bradbury, E. J.; Moon, L. D. F.; Popat, R. J.; King, V. R.; Bennett, G. S.; Patel, P. N.; Fawcett, J. W.; McMahon, S. B. *Nature* **2002**, *416*, 636.

- (3) Brown, J. M.; Xia, J.; Zhuang, B. Q.; Cho, K. S.; Rogers, C. J.; Gama, C. I.; Rawat, M.; Tully, S. E.; Uetani, N.; Mason, D. E.; Tremblay, M. L.; Peters, E. C.; Habuchi, O.; Chen, D. F.; Hsieh-Wilson, L. C. *P. Natl. Acad. Sci. U.S.A.* **2012**, *109*, 4768.
- (4) Sepulveda-Diaz, J. E.; Naini, S. M. A.; Huynh, M. B.; Ouidja, M. O.; Yanicostas, C.; Chantepie, S.; Villares, J.; Lamari, F.; Jospin, E.; van Kuppevelt, T. H.; Mensah-Nyagan, A. G.; Raisman-Vozari, R.; Soussi-Yanicostas, N.; Papy-Garcia, D. *Brain* **2015**, *138*, 1339.
- (5) Reynard, L. N.; Loughlin, J. *Nat. Rev. Rheumatol.* **2013**, *9*, 573.
- (6) Plaas, A. H.; West, L. A.; Wong-Palms, S.; Nelson, F. R. *J. Biol. Chem.* **1998**, *273*, 12642.
- (7) Sasisekharan, R.; Shriver, Z.; Venkataraman, G.; Narayanasami, U. *Nat. Rev. Cancer* **2002**, *2*, 521.
- (8) Mizumoto, S.; Watanabe, M.; Yamada, S.; Sugahara, K. *Biomed. Res. Int.* **2013**.
- (9) Fried, M.; Duffy, P. E. *Science* **1996**, *272*, 1502.
- (10) Beeson, J. G.; Chai, W. G.; Rogerson, S. J.; Lawson, A. M.; Brown, G. V. *Infect. Immun.* **1998**, *66*, 3397.
- (11) Shukla, D.; Liu, J.; Blaiklock, P.; Shworak, N. W.; Bai, X.; Esko, J. D.; Cohen, G. H.; Eisenberg, R. J.; Rosenberg, R. D.; Spear, P. G. *Cell* **1999**, *99*, 13.
- (12) Funderburgh, J. L. *Glycobiology* **2000**, *10*, 951.
- (13) Akama, T. O.; Nishida, K.; Nakayama, J.; Watanabe, H.; Ozaki, K.; Nakamura, T.; Dota, A.; Kawasaki, S.; Inoue, Y.; Maeda, N.; Yamamoto, S.; Fujiwara, T.; Thonar, E. J.; Shimomura, Y.; Kinoshita, S.; Tanigami, A.; Fukuda, M. N. *Nat. Genet.* **2000**, *26*, 237.
- (14) Burkart, M. D.; Izumi, M.; Chapman, E.; Lin, C. H.; Wong, C. H. *J. Org. Chem.* **2000**, *65*, 5565.
- (15) Chapman, E.; Ding, S.; Schultz, P. G.; Wong, C. H. *J. Am. Chem. Soc.* **2002**, *124*, 14524.
- (16) Lin, E. S.; Yang, Y. S. *Anal. Biochem.* **1998**, *264*, 111.
- (17) Chen, W. T.; Liu, M. C.; Yang, Y. S. *Anal. Biochem.* **2005**, *339*, 54.
- (18) Macarron, R.; Hertzberg, R. P. *Mol. Biotechnol.* **2011**, *47*, 270.

Chapter 3

STRUCTURE-ACTIVITY RELATIONSHIP ANALYSIS OF LEAD COMPOUNDS

Chapter 3

STRUCTURE-ACTIVITY RELATIONSHIP ANALYSIS OF LEAD SCAFFOLDS

Introduction

The high-throughput screen we previously developed revealed two lead scaffolds, **5** and **19**, as millimolar inhibitors of Chst15. Traditionally, to learn about the implications of the structural components of each scaffold, libraries of compound analogs are often assembled. Decades of medicinal chemistry research have taught us that these chemical changes, most often composed of subtle, isoteric variations, but may also consist of large molecular additions, can help elucidate the key moieties important for inhibitor potency.¹⁻
³ Here, we obtained commercially available and chemically synthesized analogs of lead compounds **5** and **19**. Using a radioactive isotope labeling assay, we evaluated the structure-activity relationship of the changes as compared to the original hit compounds. We determined that compound **19** would be the most promising scaffold to further develop based on specificity and chemical tractability of the hit molecules. Fine tuning of the substituents on the anilide moiety of **19** led to a 3-fold improvement. This optimized compound will be carried through for further cellular and potentially *in vivo* analysis.

Results and Discussion

To assess the structure-activity relationship of these two molecules, we obtained commercially available and chemically synthesized analogs based around the two pharmacophores. Several different analogs of **5** were obtained commercially and tested with the radioisotope labeling assay; however, none of the changes improved the inhibition relative to the original compound when tested at 20 μ M (Figure 3.1). Interestingly, **5** is a known c-Raf inhibitor and when retested in the counter screen, inhibited Sult1c1 as well (Figure 2.6A).⁴ Due to these potential off-target effects, we decided to focus our attention on **19**.

Thirty-one analogs of **19** were synthesized chemically or obtained from commercial sources. To generate the analogs, we used two different synthetic routes depending on the molecule's diversification point of interest. Synthesis of the α,β -unsaturated acid first gave us access to changing the functional groups around the anilide by *N,N'*-diisopropylcarbodiimide (DIC) promoted amide bond coupling (Scheme 3.1). Synthesis of the cyanoacetamide in two steps allowed us to append different aryl-aldehyde intermediates by piperidinium acetate-catalyzed Knoevenagel condensation (Scheme 3.2).⁵ Compounds were purified by column chromatography and if necessary by HPLC purification.

Using the same radioisotope labeling assay as before, we were able to compare the relative inhibitory activity of each small molecule (Table 3.1, Table 3.2, and Figure 3.2). First, we individually removed the aromatic rings on each end of the molecule and, as expected, abolished bioactivity (**21**, **22**, **25**), indicative of their importance in the

pharmacophore. We then sought to change the substituents on the anilide functional group. We reasoned that changing the substituents at the position “para” to the rest of the molecule would likely extend it most and potentially change steric interactions if located in a binding cleft. However, our findings indicated that removing the di-substitution on the anilide ring greatly decreased the relative inhibitory activity as was seen with all analogs containing only a mono-substituted ring (**28-36**). Altering the position of the mono-substitution (**34**, **35**) also did not change this trend. Even large substitutions, such as tert-butyl (**33**) and phenyl groups (**36**, **50**), were not able to compensate for the loss in activity. This was especially evident in the di-chloro analogs (**37-39**), as they all had comparable, if not greater, potency to the original lead compound **19**.

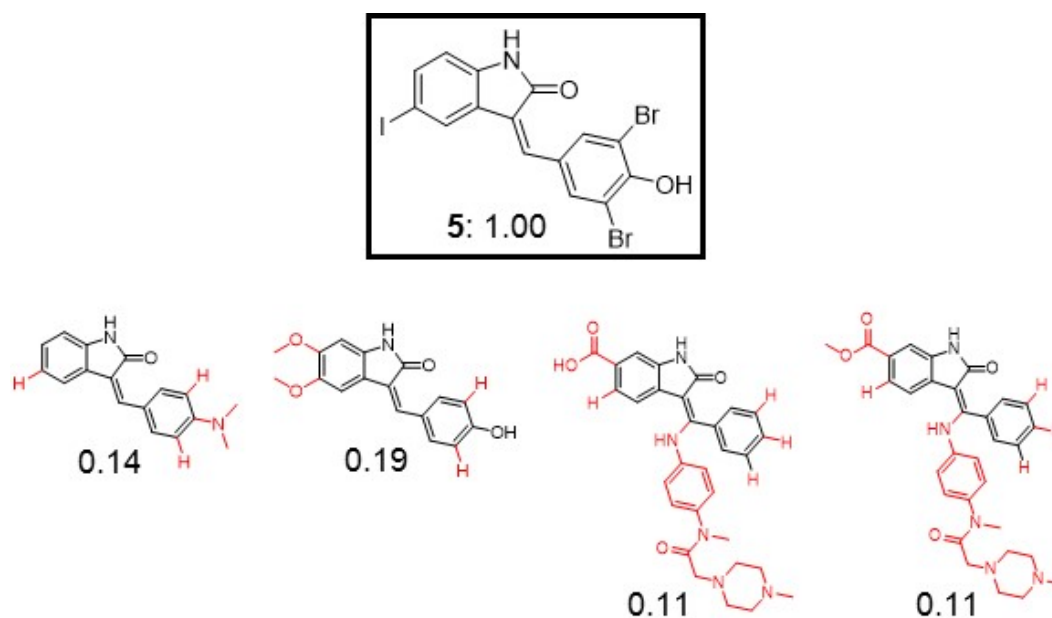
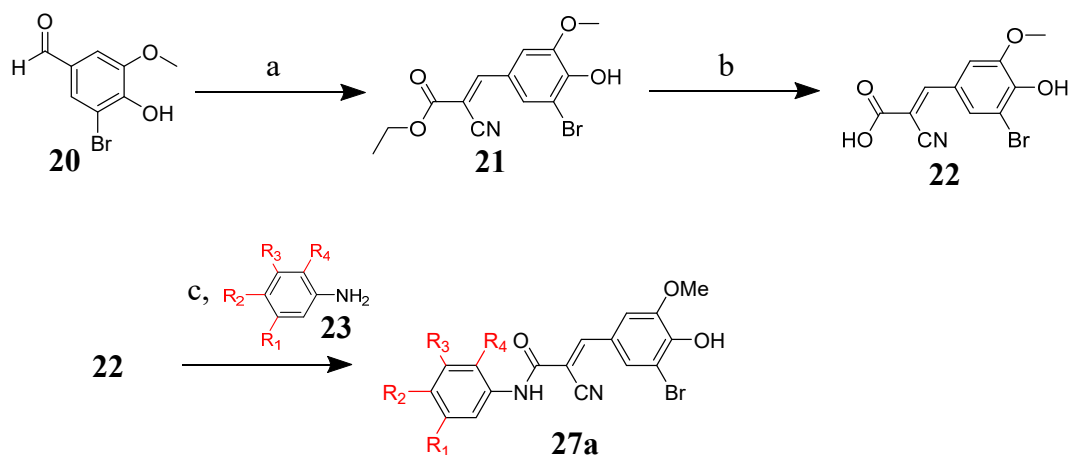
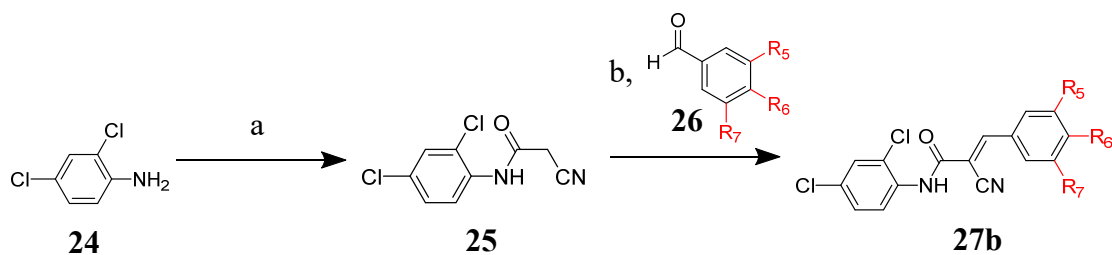


Figure 3.1. Comparison of chemical analogs to lead compound **5**. Relative inhibition values are given for each compound as compared to the original hit **5**. Compounds were assessed through the radiolabeling assay at 20 μ M. Mean values of duplicate experiments are reported.

Scheme 3.1^a.

^aReagents and conditions: (a) ethyl cyanoacetate, piperidine, AcOH, PhMe (91%); (b) NaOH, THF (77%); (c) DIC, THF (17-56%).

Scheme 3.2^a.

^aReagents and conditions: (a) cyanoacetic acid, DIC, THF (45%); (b) piperidine AcOH, PhMe (52-93%).

Removal of any of the substituents on the vinylbenzene ring (**40**, **44**, **47**) greatly attenuated the inhibitory potency of the compound indicating the importance each has in the bioactivity of the molecule. Replacement of the original bromo substituent with different halogens showed that the steric interactions elicited by the bromo group was ideal for the binding pocket (**41-43**). While the larger iodo substituent was tolerated, there was a slight decline in activity (**43**). Smaller halogens were not as well tolerated and led to a 2-2.5-fold loss in inhibitory activity (**41**, **42**) as compared to the original lead compound.

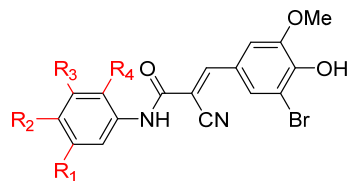
Slight modification to bromo (**45**) or extension to ethoxy (**46**) of the original methoxy group produced compounds with inhibitory activities comparable to or better than that of the original lead scaffold. However, when comparing to the most active, similar scaffold, **37**, only the ethoxy group change is well tolerated by retaining 93% of its inhibitory activity. Extension at this position could potentially be used for further elaboration in the future.

Meanwhile, amino and indazole isosteres of the phenol group (**48**, **51**) were not well tolerated, with a drop in potency of 3-fold and 26-fold respectively. Methylation of the phenol functionality also drastically lowered inhibitory potency when comparing compounds with common scaffolds (**38**, **49**). It is likely that the electronic and hydrogen bonding properties of the phenol group are necessary for the bioactivity of the inhibitor.

Interestingly, reduction of the acrylamide double bond led to a 9-fold reduction in potency and loss of dose dependent inhibitory activity (**52**). Mechanistically, this alkene could potentially serve as an electrophilic site for a nucleophilic amino acid sidechain and/or increase the rigidity of the molecule by restricting the rotational freedom of the

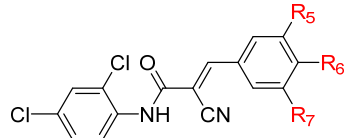
bond. Rotating frame nuclear Overhauser effect spectroscopy (ROESY) data suggested that the *E*-isomer is generated from the Knoevengel condensation (Experimental Methods). Coupling between the amide and alkene protons was detected in **37** and is only possible in the *E*-configuration.

Table 3.1. Chst15 Inhibition Relative to **19** of Anilide-Analogs at 25 μ M.



| Compound | R ₁ | R ₂ | R ₃ | R ₄ | Relative Inhibition |
|-----------|----------------|-----------------|----------------|----------------|---------------------|
| 19 | H | F | Cl | H | 1.00 \pm 0.01 |
| 28 | H | F | H | H | 0.19 \pm 0.00 |
| 29 | H | Cl | H | H | 0.30 \pm 0.04 |
| 30 | H | Br | H | H | 0.10 \pm 0.00 |
| 31 | H | Me | H | H | 0.08 \pm 0.00 |
| 32 | H | CF ₃ | H | H | 0.10 \pm 0.00 |
| 33 | H | tBu | H | H | 0.35 \pm 0.05 |
| 34 | H | H | H | Cl | 0.57 \pm 0.02 |
| 35 | H | H | H | F | 0.37 \pm 0.00 |
| 36 | H | H | Ph | H | 0.07 \pm 0.01 |
| 37 | H | Cl | H | Cl | 1.35 \pm 0.12 |
| 38 | Cl | H | Cl | H | 0.98 \pm 0.10 |
| 39 | Cl | H | H | Cl | 1.21 \pm 0.09 |

Table 3.2. Chst15 Inhibition Relative to **19** of Vinylbenzene-Analogs at 25 μ M.



| Compound | R ₅ | R ₆ | R ₇ | Relative Inhibition |
|-----------|----------------|-----------------|----------------|---------------------|
| 37 | OMe | OH | Br | 1.35 \pm 0.12 |
| 40 | OMe | OH | H | 0.13 \pm 0.00 |
| 41 | OMe | OH | F | 0.40 \pm 0.03 |
| 42 | OMe | OH | Cl | 0.49 \pm 0.01 |
| 43 | OMe | OH | I | 0.88 \pm 0.02 |
| 44 | H | OH | Br | 0.32 \pm 0.00 |
| 45 | Br | OH | Br | 0.97 \pm 0.10 |
| 46 | OEt | OH | Br | 1.26 \pm 0.16 |
| 47 | OMe | H | Br | 0.00 \pm 0.00 |
| 48 | OMe | NH ₂ | Br | 0.32 \pm 0.01 |

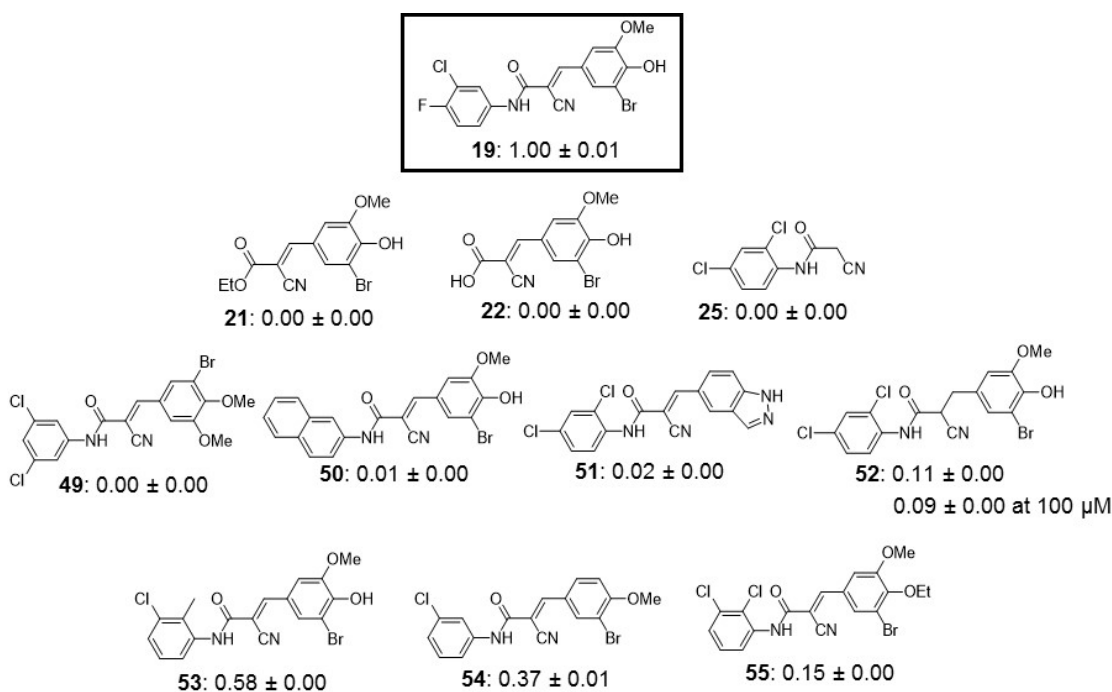


Figure 3.2. Comparison of additional chemical analogs to lead compound **19**. Relative inhibition values are given for each compound as compared to the original hit **19**. Compounds were assessed through the radiolabeling assay at 25 μ M. The mean \pm S.E.M. of triplicate experiments is reported.

Conclusion

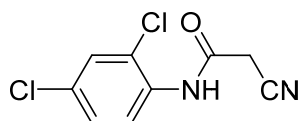
In conclusion, here we used commercially available and chemically synthesized analogs to evaluate the structure-activity relationships of the components of the lead scaffolds. Due to the known off target effects of compound **5**, we decided to focus our attention on developing **19**. A straightforward, 2-3 step synthetic route led to the facile assembly of **19** analogs. Comparative analysis revealed many structural components that were critical for inhibitory activity, including the acrylamide olefin and the vinylbenzene phenol. Subtle changes to the substituents around the anilide ring were able to potentiate a 3-fold increase in potency and this compound, **37**, will be characterized further. For future

studies, to accelerate the generation of more potent and potentially specific inhibitors, it would be ideal to gain structural information about how the molecule interacts with the target enzyme, Chst15. This would greatly narrow down the possible changes to the scaffold and facilitate rational design of the inhibitor.

Experimental Methods

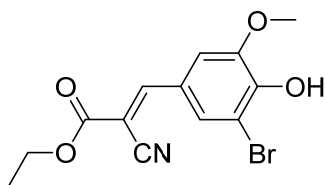
General Methods. All reactions were carried out under an argon atmosphere unless otherwise specified. Reagents were purchased from Sigma Aldrich and used without purification unless otherwise specified. ^1H NMR spectra were recorded with a Varian Inova 500 (500 MHz) and are reported relative to residual solvent peaks. ^{13}C NMR spectra were recorded with Varian Inova 500 (125 MHz) or Bruker AVANCE AV400 (100 MHz) and are reported relative to residual solvent peaks. ROESY NMR spectra were recorded with a Varian Inova 600 (600 MHz) spectrometer. Data for ^1H NMR are reported as follows: chemical shift (δ ppm) (multiplicity, coupling constant (Hz), integration, assignment). Multiplicities are reported as follows: s = singlet, d = doublet, t = triplet, m = multiplet, br = broad. Data for ^{13}C NMR are reported in terms of chemical shifts (δ ppm). High resolution mass spectra (HRMS) were acquired using an Agilent 6200 Series TOF with and Agilent G1978A Multimode source using mixed electrospray ionization/atmospheric pressure chemical ionization (MM: ESI-APCI).

2-cyano-N-(2,4-dichlorophenyl)acetamide



Cyanoacetic acid (0.263 g, 3.09 mmol) was dissolved into 10 mL of THF at room temperature. A solution of 2,4-dichloroaniline (0.500, 3.09 mmol) in 5 mL of THF was added dropwise at room temperature. After 5 minutes, a solution of DIC (0.478 mL, 3.09 mmol) in 5 mL of THF was added dropwise. The reaction was refluxed under argon for 1 hr. The flask was then allowed to return to room temperature. The mixture was then filtered and the filtrate was concentrated *in vacuo*. The crude solid was then redissolved into EtOAc (20 mL) and the organic layer was extracted with 6 N HCl (20 mL x 3) and saturated NaHCO₃ (20 mL x 3) successively. The organic layer was dried with MgSO₄ and concentrated *in vacuo*. The solid was then recrystallized in DCM to afford a white solid (0.319 g, 45%). *R*_f 0.2 (8:2 Hex/EtOAc). ¹H NMR (500 MHz, CD₃OD) δ 8.25 (d, *J* = 8.9 Hz, 1 H, *ArH*), 8.21 (br, 1 H, *NH*), 7.44 (d, *J* = 2.4 Hz, 1 H, *ArH*), 7.30 (dd, *J* = 2.4, 8.9 Hz, 1 H, *ArH*), 3.62 (s, 2 H, *CH*₂CN); ¹³C NMR (125 MHz, DMSO-*d*₆) δ 159.23, 132.46, 131.18, 129.47, 128.51, 124.68, 123.00, 114.20, 27.47. HRMS: [*M*-H]⁻ calculated: 226.9784, found: 226.9792.

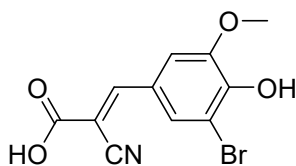
(E)-Ethyl 3-(3-bromo-4-hydroxy-5-methoxyphenyl)-2-cyanoacrylate



Ethyl cyanoacetate (1.12 mL, 10.5 mmol) and 5-bromovanillin (2.31 g, 10 mmol) was dissolved into 10 mL of toluene at room temperature. Piperidine (0.158 mL, 1.60 mmol) and glacial acetic acid (0.74 mL, 12.9 mmol) were then added to the previous mixture. The

mixture was refluxed for 3 hours. After 3 hours, the reaction was allowed to cool to room temperature. The precipitate was filtered, collected, and rinsed with ice cold methanol to yield a yellow solid (2.98 g, 91%). R_f 0.2 (1:1 Hex/EtOAc). ^1H NMR (500 MHz, DMSO- d_6) δ 8.16 (s, 1 H, C=CH), 7.79 (s, 2 H, ArH), 4.35 (q, J = 7.1 Hz, 2 H, $\text{CH}_3\text{CH}_2\text{O}$), 3.94 (s, 3 H, OCH_3), 1.37 (t, J = 7.2 Hz, 3 H, $\text{CH}_3\text{CH}_2\text{O}$); ^{13}C NMR (125 MHz, DMSO- d_6) δ 162.62, 154.03, 149.62, 148.64, 129.49, 123.90, 116.61, 113.56, 110.15, 99.54, 62.57, 56.78, 14.47. HRMS: $[\text{M}-\text{H}]^-$ calculated: 323.9877, found: 323.9883.

(E)-3-(3-bromo-4-hydroxy-5-methoxyphenyl)-2-cyanoacrylic acid

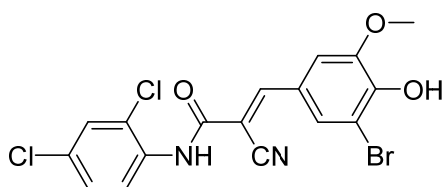


Previously synthesized ethyl 3-(3-bromo-4-hydroxy-5-methoxyphenyl)-2-cyanoacrylate (0.101 g, 0.311 mmol) was dissolved in 4 mL of THF. 2 N NaOH (10 mL) was added. The mixture was stirred at room temperature for 15 hours. After 15 hours, the reaction was neutralized with concentrated HCl. The reaction mixture was then extracted with EtOAc (3 x 10 mL). The organic layer was then washed with brine (10 mL), dried with MgSO_4 , and concentrated *in vacuo*. The resulting crude solid was then redissolved into 10:1 DCM/EtOAc and purified by silica flash chromatography using 10:1 DCM/EtOAc and 0.1% AcOH. A yellow solid was obtained after concentration *in vacuo* (0.071 g, 77%). R_f 0.3 (9:1 DCM/MeOH). ^1H NMR (500 MHz, DMSO- d_6) δ 8.21 (s, 1 H, C=CH), 7.93 (s, 2 H, ArH), 7.77 (s, 1 H, ArH), 3.88 (s, 3H, OCH_3), 3.12 (s, 1 H, ArOH); ^{13}C NMR (125

MHz, DMSO- d_6) δ 165.43, 163.86, 151.12, 148.37, 128.51, 123.15, 117.65, 116.11, 112.45, 109.96, 56.16. HRMS: $[M-H]^-$ calculated: 295.9564, found: 295.9578.

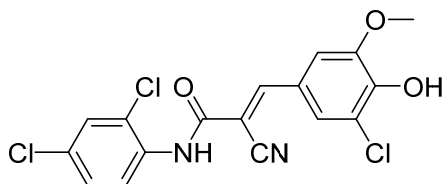
General Procedure for Aldol Reaction:

(E)-3-(3-bromo-4-hydroxy-5-methoxyphenyl)-2-cyano-N-(2,4-dichlorophenyl)acrylamide



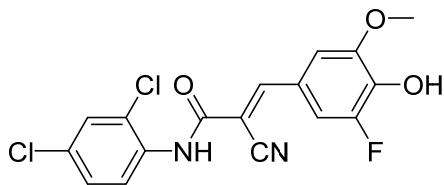
Previously synthesized 2-cyano-N-(2,4-dichlorophenyl)acetamide (0.240 g, 1.05 mmol) and 5-bromovanillin (0.231 g, 1 mmol) was added to 5 mL of toluene. While stirring this mixture, piperidine (0.016 mL, 0.16 eq) and glacial acetic acid (0.074 mL, 1.25 eq) was added. The solution was stirred and heated to reflux for 3 hours. After 3 hours, the reaction was allowed to cool to room temperature. The precipitate was filtered, collected, and rinsed with ice cold methanol to yield a yellow solid (0.413 g, 93%). 1H NMR (500 MHz, DMSO- d_6) δ 10.88 (s, 1 H, ArOH), 10.01 (s, 1 H, NH), 8.22 (s, 1 H, C=CH), 7.90 (d, J = 2.0 Hz, 1 H, ArH), 7.76 (d, J = 2.4 Hz, 1 H, ArH), 7.73 (d, J = 2.0 Hz, 1 H, ArH), 7.66 (d, J = 8.6 Hz, 1 H, ArH), 7.49 (dd, J = 2.4, 8.6 Hz, 1 H, ArH), 3.89 (s, 3 H, OCH_3); ^{13}C NMR (125 MHz, DMSO- d_6) δ 160.58, 150.93, 148.62, 133.63, 131.00, 129.82, 129.15, 128.56, 128.32, 127.83, 123.70, 116.73, 112.62, 109.34, 102.24, 56.33. HRMS: $[M-H]^-$ calculated: 438.9257, found: 438.9258.

(E)-3-(3-chloro-4-hydroxy-5-methoxyphenyl)-2-cyano-*N*-(2,4-dichlorophenyl)acrylamide



3-Chloro-4-hydroxy-5-methoxybenzaldehyde (0.078 g, 0.417 mmol) was used in the same previous aldol reaction conditions (0.151 g, 91%); ^1H NMR (500 MHz, DMSO- d_6) δ 10.83 (s, 1 H, ArOH), 9.95 (s, 1 H, NH), 8.20 (s, 1 H, C=CH), 7.75 (d, J = 2.2 Hz, 2 H, ArH), 7.68 (d, J = 8.6 Hz, 2 H, ArH), 7.49 (dd, J = 2.4, 8.6 Hz, 1 H, ArH), 3.88 (s, 3 H, OCH_3); ^{13}C NMR (125 MHz, DMSO- d_6) δ 161.00, 151.47, 149.09, 148.22, 134.08, 131.40, 130.18, 129.57, 128.91, 128.26, 125.82, 123.29, 120.86, 117.18, 112.67, 102.73, 56.77. HRMS: $[\text{M}-\text{H}]^-$ calculated: 394.9762, found: 394.9780.

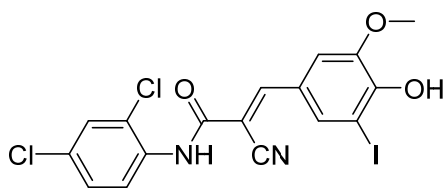
(E)-2-cyano-*N*-(2,4-dichlorophenyl)-3-(3-fluoro-4-hydroxy-5-methoxyphenyl)acrylamide



3-fluoro-4-hydroxy-5-methoxybenzaldehyde (AstaTech Inc, 0.035 g, 0.208 mmol) was used in the same previous aldol reaction conditions (0.068 g, 86%); ^1H NMR (500 MHz, DMSO- d_6) δ 9.94 (s, 1 H, NH), 8.20 (s, 1 H, C=CH), 7.74 (d, J = 2.4 Hz, 1 H, ArH), 7.68 (d, J = 8.7 Hz, 1 H, ArH), 7.58 (m, 2 H, ArH), 7.48 (dd, J = 2.4, 8.7 Hz, 1 H, ArH), 3.86 (s,

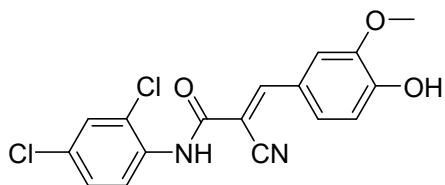
3 H, OCH_3); ^{13}C NMR (125 MHz, DMSO-d_6) δ 161.16, 152.56, 151.64, 150.66, 150.24, 150.19, 134.18, 131.20, 129.89, 129.53, 128.60, 128.25, 117.49, 112.29, 112.13, 111.19, 56.76. HRMS: $[\text{M-H}]^-$ calculated: 379.0058, found: 379.0074.

(E)-2-cyano-N-(2,4-dichlorophenyl)-3-(4-hydroxy-3-iodo-5-methoxyphenyl)acrylamide



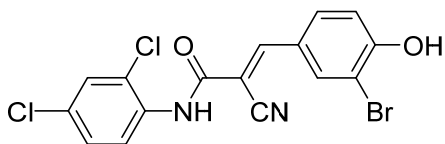
5-iodovanillin (0.058g, 0.208 mmol) was used in the same previous aldol reaction conditions (0.072 g, 71%).; ^1H NMR (500 MHz, DMSO-d_6) δ 10.87 (s, 1 H, NH), 9.90 (br, 1 H, OH), 8.16 (s, 1 H, C=CH), 8.04 (s, 1 H, ArH), 7.74 (d, $J = 2.4$ Hz, 1 H, ArH), 7.72 (s, 1 H, ArH), 7.68 (d, $J = 8.7$ Hz, 1 H, ArH), 7.48 (dd, $J = 2.4, 8.6$ Hz, 1 H, ArH), 3.86 (s, 3 H, OCH_3); ^{13}C NMR (125 MHz, DMSO-d_6) δ 160.54, 151.21, 150.53, 146.81, 134.42, 133.56, 130.78, 129.52, 129.00, 127.69, 124.60, 116.67, 113.06, 101.77, 84.94, 56.11. HRMS: $[\text{M-H}]^-$ calculated: 486.9119, found: 486.9132.

(E)-2-cyano-N-(2,4-dichlorophenyl)-3-(4-hydroxy-3-methoxyphenyl)acrylamide



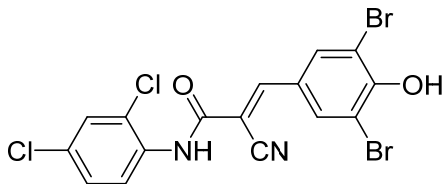
Vanillin (0.032 g, 0.208 mmol) was used in the same previous aldol reaction conditions (0.067 g, 89%); ^1H NMR (500 MHz, DMSO- d_6) δ 10.40 (s, 1 H, ArOH), 9.92 (s, 1 H, NH), 8.23 (s, 1 H, C=CH), 7.73 (m, 2 H, ArH), 7.67 (d, J = 8.6 Hz, 1 H, ArH), 7.57 (d, J = 2.1, 8.4 Hz, 1 H, ArH) 7.48 (dd, J = 2.4, 8.7 Hz, 1 H, ArH), 6.98 (d, J = 8.3 Hz, 1 H, ArH), 3.86 (s, 3 H, OCH₃); ^{13}C NMR (125 MHz, DMSO- d_6) δ 160.78, 152.06, 152.00, 147.76, 133.64, 130.67, 129.45, 128.96, 128.17, 127.66, 126.29, 123.00, 117.00, 115.99, 113.67, 100.45, 55.58. HRMS: $[\text{M-H}]^-$ calculated: 361.0152, found: 361.0172.

(E)-3-(3-bromo-4-hydroxyphenyl)-2-cyano-N-(2,4-dichlorophenyl)acrylamide



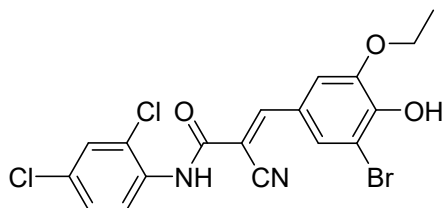
3-Bromo-4-hydroxybenzaldehyde (0.042 g, 0.208 mmol) was used in the same previous aldol reaction conditions (0.044 g, 52%); ^1H NMR (500 MHz, DMSO- d_6) δ 9.96 (s, 1 H, NH), 8.26 (d, J = 2.2 Hz, 1 H, ArH), 8.19 (s, 1 H, C=CH), 7.93 (dd, J = 2.3, 8.7 Hz, 1 H, ArH), 7.74 (d, J = 2.4 Hz, 1 H, ArH), 7.66 (d, J = 8.7 Hz, 1 H, ArH) 7.48 (dd, J = 2.4, 8.6 Hz, 1 H, ArH), 7.10 (d, J = 8.6 Hz, 1 H, ArH); ^{13}C NMR (125 MHz, DMSO- d_6) δ 161.11, 159.51, 159.48, 150.85, 136.11, 134.06, 132.45, 131.32, 130.14, 129.54, 128.90, 128.23, 124.24, 117.40, 117.13, 110.86. HRMS: $[\text{M-H}]^-$ calculated: 408.9152, found: 408.9166.

(E)-2-cyano-3-(3,5-dibromo-4-hydroxyphenyl)-N-(2,4-dichlorophenyl)acrylamide

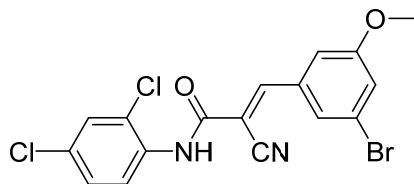


3,5-Dibromo-4-hydroxybenzaldehyde (0.058 g, 0.208 mmol) was used in the same previous aldol reaction conditions (0.084 g, 82%); ^1H NMR (500 MHz, DMSO- d_6) δ 10.03 (s, 1 H, *NH*), 8.27 (s, 2 H, *ArH*), 8.20 (s, 1 H, *C=CH*), 7.75 (d, $J = 2.4$ Hz, 1 H, *ArH*), 7.67 (d, $J = 8.7$ Hz, 1 H, *ArH*), 7.49 (dd, $J = 2.4, 8.6$ Hz, 1 H, *ArH*); ^{13}C NMR (125 MHz, DMSO- d_6) δ 159.98, 154.88, 148.72, 134.22, 133.24, 130.69, 129.32, 128.83, 128.05, 127.52, 125.08, 115.92, 111.84, 103.44. HRMS: $[\text{M}-\text{H}]^-$ calculated: 486.8257, found: 486.8264.

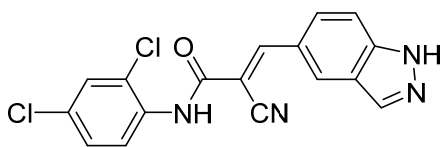
(E)-3-(3-bromo-5-ethoxy-4-hydroxyphenyl)-2-cyano-N-(2,4-dichlorophenyl)acrylamide



3-Bromo-5-ethoxy-4-hydroxybenzaldehyde (Alfa Aesar, 0.051 g, 0.208 mmol) was used in the same previous aldol reaction conditions (0.081 g, 85%); ^1H NMR (500 MHz, DMSO- d_6) δ 10.62 (br s, 1 H, *OH*), 10.00 (s, 1 H, *NH*), 8.21 (s, 1 H, *C=CH*), 7.87 (d, $J = 1.7$ Hz, 1 H, *ArH*), 7.75 (d, $J = 2.3$ Hz, 1 H, *ArH*), 7.72 (d, $J = 1.7$ Hz, 1 H, *ArH*), 7.65 (d, $J = 8.7$ Hz, 1 H, *ArH*), 7.49 (dd, $J = 2.3, 8.6$ Hz, 1 H, *ArH*); ^{13}C NMR (125 MHz, DMSO- d_6) δ 160.93, 151.34, 149.26, 147.81, 134.00, 131.32, 129.86, 129.49, 128.80, 128.56, 128.21, 124.25, 117.08, 113.84, 110.16, 102.89, 65.31, 14.79. HRMS: $[\text{M}-\text{H}]^-$ calculated: 452.9414, found: 452.9429.

(E)-3-(3-bromo-5-methoxyphenyl)-2-cyano-*N*-(2,4-dichlorophenyl)acrylamide

3-Bromo-5-methoxybenzaldehyde (AstaTech Inc., 0.100 g, 0.465 mmol) was used in the same previous aldol reaction conditions (0.138 g, 70%); ^1H NMR (500 MHz, DMSO- d_6) δ 10.23 (s, 1 H, NH), 8.30 (s, 1 H, C=CH), 7.76 (d, J = 1.6 Hz, 1 H, ArH), 7.75 (d, J = 2.4 Hz, 1 H, ArH), 7.63 (d, J = 8.6 Hz, 1 H, ArH), 7.58 (t, J = 1.7 Hz, 1 H, ArH), 7.49 (dd, J = 2.4, 8.6 Hz, 1 H, ArH), 7.43 (t, J = 1.6 Hz, 1 H, ArH), 3.84 (s, 3 H, OCH_3); ^{13}C NMR (126 MHz, DMSO- d_6) δ 160.76, 160.45, 150.82, 134.99, 133.80, 131.72, 130.45, 129.66, 129.22, 128.32, 125.07, 123.25, 121.03, 116.15, 115.42, 108.11, 56.39. HRMS: $[\text{M}-\text{H}]^-$ calculated: 422.9308, found: 422.9204.

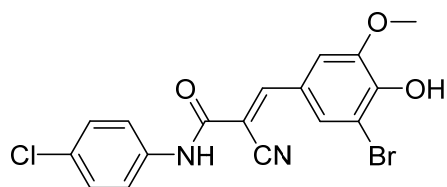
(E)-2-cyano-*N*-(2,4-dichlorophenyl)-3-(1*H*-indazol-5-yl)acrylamide

Indazole-5-carboxaldehyde (0.016 g, 0.109 mmol) was used in the same previous aldol reaction conditions (0.024 g, 62%); ^1H NMR (500 MHz, DMSO- d_6) 13.53 (s, 1 H, ArNH), 10.07 (s, 1 H, NH), 8.51 (s, 1 H, ArH), 8.45 (s, 1 H, C=CH), 8.32 (s, 1 H, ArH), 8.09 (dd, J = 1.7, 8.9 Hz, 1 H, ArH), 7.75 (m, 2 H, ArH), 7.68 (d, J = 8.7 Hz, 1 H, ArH), 7.49 (dd, J = 2.4, 8.6 Hz, 1 H, ArH); ^{13}C NMR (126 MHz, DMSO- d_6) δ 161.26, 153.41, 136.12,

134.11, 131.44, 130.32, 129.59, 129.09, 128.28, 127.32, 126.90, 124.72, 123.63, 117.32, 111.74, 109.99, 102.96. HRMS: $[M-H]^-$ calculated: 355.0159, found: 355.0169.

General Procedure for Amide Coupling:

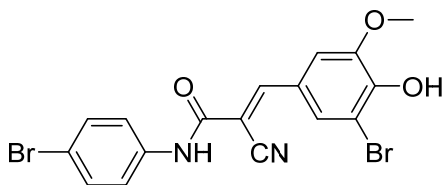
(E)-3-(3-bromo-4-hydroxy-5-methoxyphenyl)-N-(4-chlorophenyl)-2-cyanoacrylamide



Previously synthesized (E)-3-(3-bromo-4-hydroxy-5-methoxyphenyl)-2-cyanoacrylic acid (0.016 g, 0.053 mmol) and 4-chloroaniline (0.068 g, 10 eq) was dissolved in 0.5 mL of THF. *N,N'*-diisopropylcarbodiimide (DIC, 8.2 μ L, 1 eq) was added slowly and the solution was stirred at room temperature for 2.5 hours. Afterwards, the reaction mixture was diluted with 10 mL of DCM and the organic layer was extracted with 6 N HCl (3 x 10 mL), saturated NaHCO_3 (3 x 10 mL), and brine (1 x 10 mL) sequentially. The organic layer was then dried with MgSO_4 and concentrated *in vacuo*. The resulting crude solid was then redissolved into DCM with 0.1% AcOH and purified by silica flash chromatography using DCM and 0.1% AcOH. A yellow solid was obtained after concentration *in vacuo* (5 mg, 23%). R_f 0.8 (100:0.01 DCM/AcOH). ^1H NMR (500 MHz, DMSO-d_6) δ 10.82 (br, 1 H, OH), 10.41 (s, 1 H, NH), 8.15 (s, 1 H, C=CH), 7.85 (d, J = 1.9 Hz, 1 H, ArH), 7.70 (m, 3 H, ArH), 7.43 (m, 2 H, ArH), 3.89 (s, 1 H, OCH_3); ^{13}C NMR (125 MHz, DMSO-d_6) δ

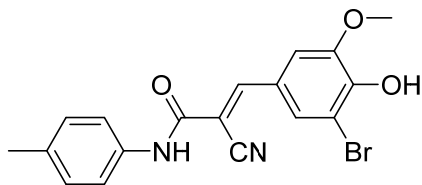
160.83, 149.95, 148.31, 148.23, 137.31, 128.68, 128.15, 127.94, 123.82, 122.11, 116.67, 112.30, 109.63, 103.76, 56.30. HRMS: $[M-H]^-$ calculated: 404.9647, found: 404.9661.

(E)-3-(3-bromo-4-hydroxy-5-methoxyphenyl)-N-(4-bromophenyl)-2-cyanoacrylamide



4-bromoaniline (0.288 g, 10 eq, 1.68 mmol) was used in the same previous amide coupling reaction conditions (0.020 g, 26%); ^1H NMR (500 MHz, $\text{CD}_3\text{CN}-d_3$) δ 8.58 (s, 1 H, *NH*), 8.10 (s, 1 H, *C=CH*), 7.78 (d, $J = 2.0$ Hz, 1 H, *ArH*), 7.70 (d, $J = 2.0$ Hz, 1 H, *ArH*), 7.59 (m, 2 H, *ArH*), 7.52 (m, 2 H, *ArH*), 3.93 (s, 3 H, OCH_3); ^{13}C NMR (125 MHz, $\text{DMSO}-d_6$) δ 160.86, 149.70, 149.43, 148.38, 137.69, 131.41, 128.42, 122.93, 122.38, 116.71, 115.86, 112.20, 109.88, 102.62, 56.17. HRMS: $[M-H]^-$ calculated: 448.9142, found: 448.9157

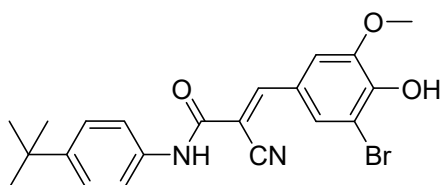
(E)-3-(3-bromo-4-hydroxy-5-methoxyphenyl)-2-cyano-N-(p-tolyl)acrylamide



p-Toluidine (0.180 g, 10 eq, 1.68 mmol) was used in the same previous amide coupling reaction conditions (0.023 g, 36%); ^1H NMR (500 MHz, CD_3OD) δ 8.03 (s, 1 H, *C=CH*), 7.73 (s, 2 H, *ArH*), 7.49 (m, 2 H, *ArH*), 7.17 (m, 2 H, *ArH*), 3.91 (s, 3 H, OCH_3), 2.33 (s,

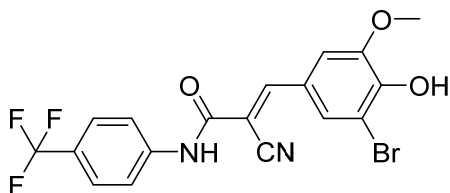
3 H, ArCH₃); ¹³C NMR (125 MHz, DMSO-d₆) δ 161.01, 152.51, 149.13, 148.86, 135.92, 132.91, 129.31, 128.92, 120.70, 120.46, 117.53, 111.83, 110.74, 99.88, 55.96, 20.37. HRMS: [M-H]⁻ calculated: 385.0193, found: 385.0208.

(E)-3-(3-bromo-4-hydroxy-5-methoxyphenyl)-N-(4-(tert-butyl)phenyl)-2-cyanoacrylamide



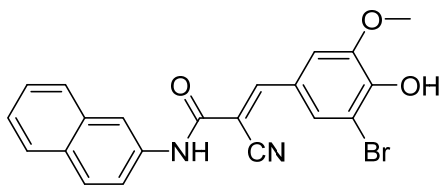
4-tert-butylaniline (0.268 mL, 10 eq, 1.68 mmol) was used in the same previous amide coupling reaction conditions (0.040 g, 56%); ¹H NMR (500 MHz, DMSO-d₆) δ 10.14 (s, 1 H, NH), 8.10 (s, 1 H, C=CH), 7.81 (d, *J* = 1.9 Hz, 1 H, ArH), 7.69 (d, *J* = 2.0 Hz, 1 H, ArH), 7.54 (m, 2 H, ArH), 7.33 (m, 2 H, ArH), 3.86 (s, 3 H, OCH₃), 1.23 (s, 9 H, C(CH₃)₃); ¹³C NMR (125 MHz, DMSO-d₆) δ 149.36, 148.19, 148.08, 146.60, 135.60, 127.90, 125.19, 123.86, 120.31, 116.62, 112.28, 109.53, 104.10, 56.24, 33.98, 31.06. HRMS: [M-H]⁻ calculated: 427.0663, found: 427.0688.

(E)-3-(3-bromo-4-hydroxy-5-methoxyphenyl)-2-cyano-N-(4-(trifluoromethyl)phenyl)acrylamide



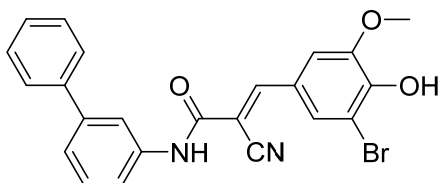
4-(Trifluoromethyl)aniline (0.211 mL, 10 eq, 1.68 mmol) was used in the same previous amide coupling reaction conditions (0.033 g, 45%); ^1H NMR (500 MHz, DMSO- d_6) δ 10.61 (s, 1 H, *NH*), 8.19 (s, 1 H, *C=CH*), 7.90 (d, J = 8.5 Hz, 2 H, *ArH*), 7.86 (d, J = 1.9 Hz, 1 H, *ArH*), 7.74 (d, J = 8.7 Hz, 2 H, *ArH*), 7.72 (d, J = 2.0 Hz, 2 H, *ArH*), 3.90 (s, 3 H, *OCH*₃); ^{13}C NMR (125 MHz, DMSO- d_6) δ 171.76, 161.14, 150.12, 148.58, 148.25, 141.90, 128.24, 125.89, 125.86, 123.55, 120.35, 116.46, 112.31, 109.63, 103.43, 56.24, 20.89. HRMS: $[\text{M}-\text{H}]^-$ calculated: 438.9911, found: 438.9950.

(E)-3-(3-bromo-4-hydroxy-5-methoxyphenyl)-2-cyano-N-(naphthalen-2-yl)acrylamide



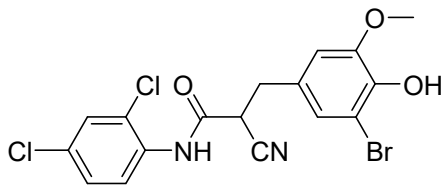
2-Naphthylamine (0.120 g, 10 eq, 0.840 mmol) was used in the same previous amide coupling reaction conditions (0.006 g, 17%); ^1H NMR (500 MHz, DMSO- d_6) 10.81 (br s, 1 H, *OH*), 10.49 (s, 1 H, *NH*), 8.30 (d, J = 2.0 Hz, 1 H, *ArH*), 8.20 (s, 1 H, *C=CH*), 7.91 (d, J = 9.0 Hz, 1 H, *ArH*), 7.87 (m, 3 H, *ArH*), 7.73 (m, 2 H, *ArH*), 7.47 (m, 2H, *ArH*), 3.90 (s, 3 H, *OCH*₃); ^{13}C NMR (101 MHz, DMSO- d_6) δ 161.02, 149.75, 148.42, 148.28, 135.99, 133.20, 130.22, 128.38, 128.21, 127.53, 127.51, 126.56, 125.15, 123.80, 120.81, 117.03, 116.82, 112.25, 109.70, 103.90, 56.31. HRMS: $[\text{M}-\text{H}]^-$ calculated: 421.0193, found: 421.0121.

(E)-N-([1,1'-biphenyl]-3-yl)-3-(3-bromo-4-hydroxy-5-methoxyphenyl)-2-cyanoacrylamide



3-Aminobiphenyl (0.142 g, 10 eq, 0.840 mmol) was used in the same previous amide coupling reaction conditions (0.017 g, 44%); ^1H NMR (500 MHz, DMSO- d_6) 10.37 (s, 1 H, NH), 8.19 (s, 1 H, C=CH), 7.98 (t, $J = 1.8$ Hz, 1 H, ArH), 7.87 (d, $J = 2.0$ Hz, 1 H, ArH), 7.73 (d, $J = 2.1$ Hz, 1 H, ArH), 7.70 (dt, $J = 1.9, 7.5$ Hz, 1 H, ArH), 7.65 (m, 2 H, ArH), 7.50 (m, 2 H, ArH), 7.46 (m, 2 H, ArH), 7.40 (m, 1 H, ArH), 3.90 (s, 3 H, OCH_3); ^{13}C NMR (101 MHz, DMSO- d_6) δ 161.27, 150.28, 148.75, 148.69, 141.17, 140.36, 139.36, 129.85, 129.49, 128.60, 128.15, 127.09, 124.30, 123.07, 119.99, 119.32, 117.21, 112.71, 110.10, 104.35, 56.75. HRMS: $[\text{M}-\text{H}]^-$ calculated: 447.0350, found: 447.0253.

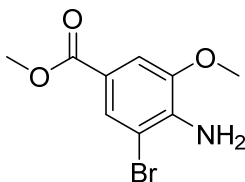
3-(3-bromo-4-hydroxy-5-methoxyphenyl)-2-cyano-N-(2,4-dichlorophenyl)propanamide



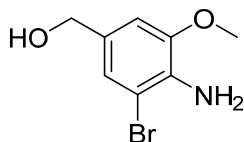
Palladium on carbon (5 mg, 10% wt % loading) was added to a round bottom flask and purged with argon. Previously synthesized (*E*)-3-(3-bromo-4-hydroxy-5-methoxyphenyl)-2-cyano-*N*-(2,4-dichlorophenyl)acrylamide (0.050 g, 0.113 mmol) was dissolved into a

minimal amount of dry acetone and added to the flask containing Pd/C. A balloon filled with H₂ was added and the reaction was stirred for 18 hours at room temperature. The mixture was filtered through Celite to remove Pd/C. The mixture was purified by preparative TLC using 4:6 EtOAc/hexanes (*R_f* 0.2) (0.005 g, 10%). ¹H NMR (500 MHz, CD₃CN) δ 8.34 (s, 1 H, *NH*), 7.87 (d, *J* = 8.8 Hz, 1 H, *ArH*), 7.53 (d, 1 H, *ArH*), 7.35 (dd, *J* = 2.4, 8.8 Hz, 1 H, *ArH*), 7.05 (d, *J* = 1.8 Hz, 1 H, *ArH*), 6.91 (d, *J* = 1.8 Hz, 1 H, *ArH*), 4.10 (t, *J* = 7.5 Hz, 1 H, *CHCN*), 3.82 (s, 1 H, *OCH*₃), 3.20 (m, 2 H, *ArCH*₂); ¹³C NMR (100 MHz, CD₃CN) δ 164.89, 148.86, 144.00, 133.81, 131.63, 130.11, 129.57, 128.66, 127.74, 126.52, 125.94, 112.78, 108.77, 57.09, 42.07, 36.41. HRMS: [M-H]⁻ calculated: 440.9414, found: 440.9434.

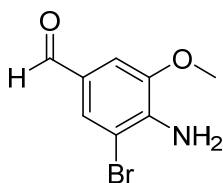
Methyl 4-amino-3-bromo-5-methoxybenzoate



This compound was prepared as previously reported. (Chem-Eur. J. 17(49), 13665-13669, 2011.) Methyl 4-amino-3-methoxybenzoate (0.500 g, 2.76 mmol) was used as previously described (0.710 g, 99%).; ¹H NMR (500 MHz, CDCl₃) δ 7.81 (d, *J* = 1.7 Hz, 1 H, *ArH*), 7.38 (d, *J* = 1.6 Hz, 1 H, *ArH*), 4.65 (s, 2 H, *NH*₂), 3.91 (s, 3 H, *COOCH*₃), 3.87 (s, 3 H, *OCH*₃); HRMS: [M+H]⁺ calculated: 259.9917, found: 259.9916.

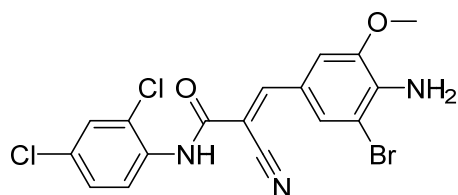
(4-amino-3-bromo-5-methoxyphenyl)methanol

LiAlH₄ (0.052 g, 1.37 mmol, 2 eq) was added to a round bottom flask under argon and placed into an ice bath. Dry diethyl ether (1.5 mL) was slowly added to the flask with stirring. A solution of the previously synthesized methyl 4-amino-3-bromo-5-methoxybenzoate (0.179 g, 0.688 mmol) and diethyl ether (1.5 mL) was added dropwise to the flask over 15 minutes. The ice bath was then removed and the reaction was allowed to stir for 2 hours at room temperature. After 2 hours, the reaction was quenched by the slow addition of EtOAc (1.5 mL). The reaction mixture was then poured into water (9 mL) and extracted with three portions of EtOAc (15 mL). The combined organic layers were dried with MgSO₄ and concentrated *in vacuo*. The resulting crude solid was then redissolved in 4:6 EtOAc/Hex and purified by silica flash chromatography using 6:4 EtOAc/Hex. The resultant compound was a clear, yellow oil (0.134 g, 84%). R_f 0.6 (6:4 EtOAc/Hex). ¹H NMR (500 MHz, CDCl₃) δ 6.93 (d, *J* = 1.8 Hz, 1 H, ArH), 6.67 (d, *J* = 1.7 Hz, 1 H, ArH), 4.45 (s, 2 H, CH₂OH), 4.14 (s, 2 H, NH₂), 3.79 (s, 3 H, OCH₃), 2.92 (br, 1 H, OH); ¹³C NMR (125 MHz, CDCl₃) δ 147.62, 133.93, 131.56, 122.90, 108.40, 108.31, 64.73, 55.94. HRMS: [M-H]⁻ calculated: 229.9822, found: 229.9821.

4-amino-3-bromo-5-methoxybenzaldehyde

Previously synthesized (4-amino-3-bromo-5-methoxyphenyl)methanol (0.131 g, 0.578 mmol) was dissolved into dry DMF (3.39 mL). MnO₂ (0.277 g, 5 eq) was added and the reaction mixture was stirred for 50 hours at room temperature. The reaction mixture was then filtered through Celite and rinsed with EtOAc (10 mL). The reaction mixture was then extracted with water (3 x 50 mL) to remove DMF. The organic layer was then dried with MgSO₄ and concentrated *in vacuo*. The resultant compound was a tan solid (0.116 g, 88%).
¹H NMR (500 MHz, CDCl₃) δ 9.64 (s, 1 H, CHO), 7.50 (d, *J* = 1.5 Hz, 1 H, ArH), 7.21 (s, *J* = 1.4 Hz, 1 H, ArH), 4.88 (s, 2 H, NH₂), 3.90 (s, 3 H, OCH₃); ¹³C NMR (125 MHz, CDCl₃) δ 189.34, 146.94, 141.20, 130.25, 127.24, 107.13, 106.31, 56.14. HRMS: [M+H]⁺ calculated: 229.9811, found: 229.9812.

(E)-3-(4-amino-3-bromo-5-methoxyphenyl)-2-cyano-N-(2,4-dichlorophenyl)acrylamide

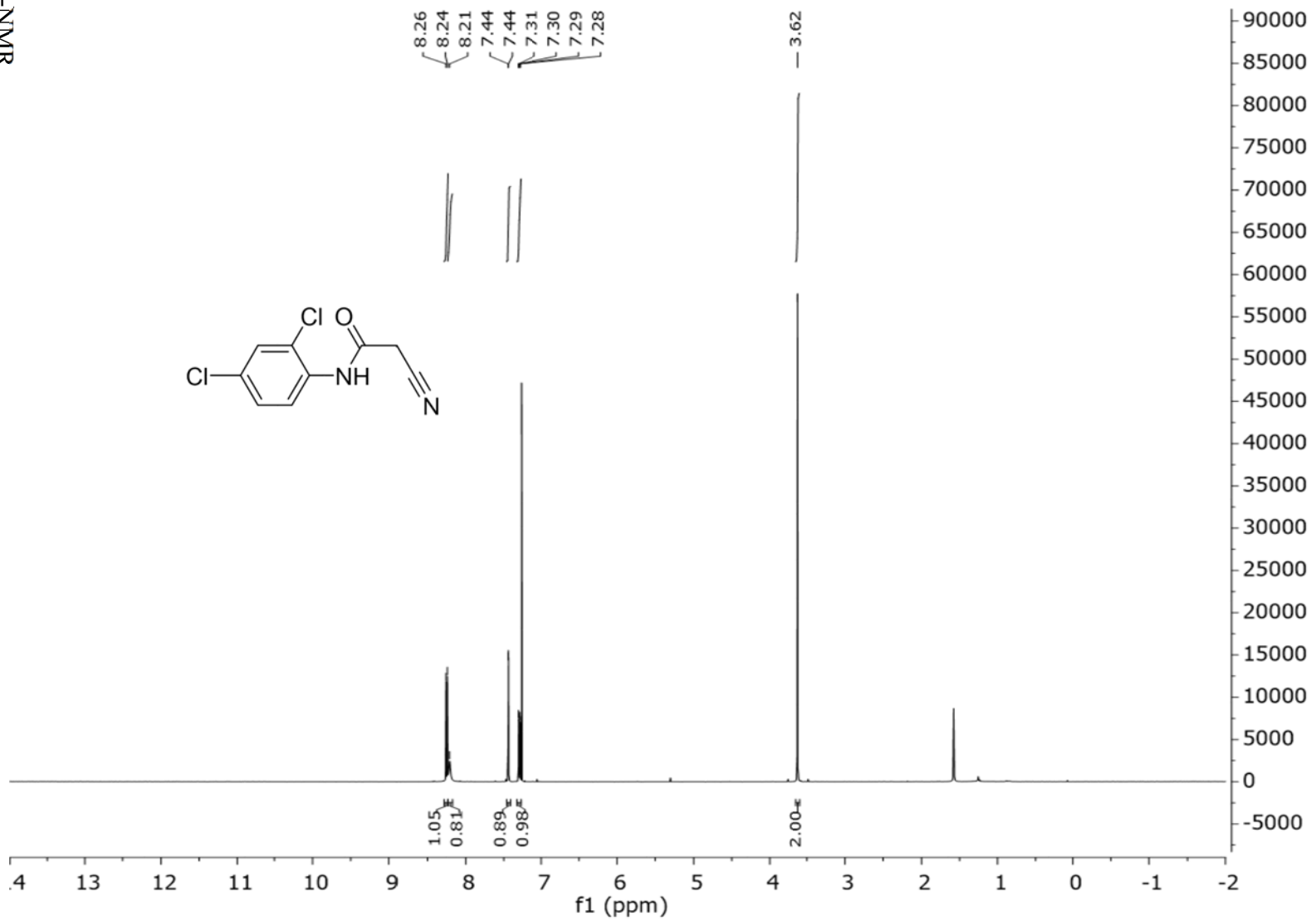


Previously synthesized 4-amino-3-bromo-5-methoxybenzaldehyde (0.048 g, 0.208 mmol) was used in the same previous aldol reaction conditions (0.081 g, 88%).; ¹H NMR (500 MHz, DMSO-d₆) δ 9.77 (s, 1 H, NH), 8.13 (s, 1 H, C=CH), 7.84 (d, *J* = 1.7 Hz, 1 H, ArH), 7.73 (d, *J* = 2.4 Hz, 1 H, ArH), 7.68 (d, *J* = 8.6 Hz, 1 H, ArH), 7.63 (d, *J* = 1.5 Hz, 1 H, ArH), 7.47 (dd, *J* = 2.4, 8.7 Hz, 1 H, ArH), 6.25 (s, 1 H, NH₂), 3.87 (s, 3 H, OCH₃); ¹³C NMR (126 MHz, DMSO-d₆) δ 161.45, 151.62, 146.42, 141.74, 134.25, 131.09, 130.16,

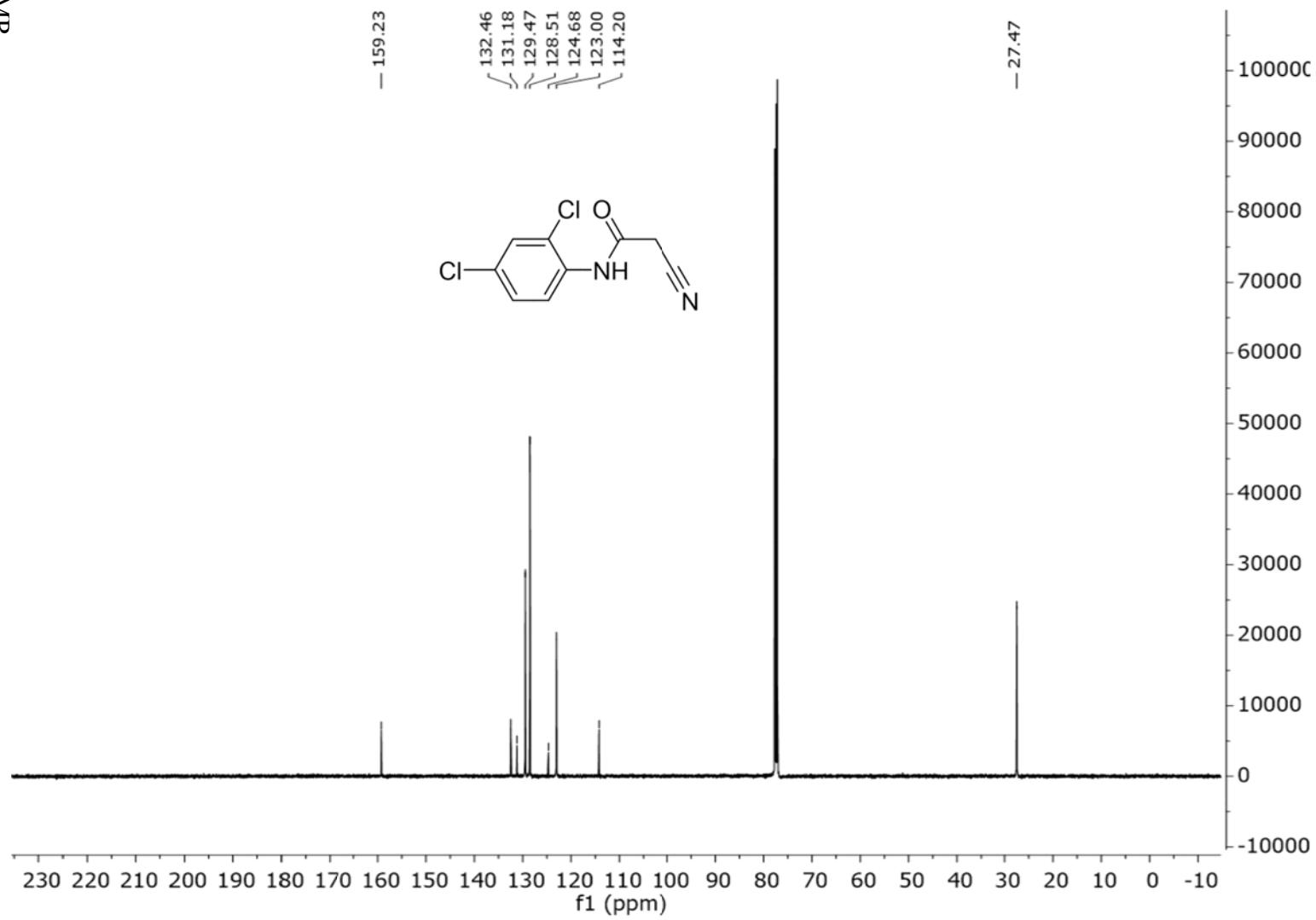
129.83, 129.47, 128.55, 128.19, 120.31, 117.93, 110.97, 105.80, 98.83, 56.49. HRMS: $[M-H]^-$ calculated: 437.9417, found: 437.9300.

Small molecule HPLC Purification. Crude small molecule was dissolved into MeOH until saturation. Purification was achieved by using an Eclipse XDB-C18 column (9.4 x 250 mm, Agilent) at a 5 mL/min flow rate with the following method on an Agilent 1100 Series HPLC: Linear gradient of 70:30 MeCN:H₂O to 88:12 MeCN:H₂O for 8 minutes, followed by 100% MeCN for 2 minutes. The compound elution was monitored by absorbance at 254 nm. Collected fractions were pooled, flash frozen, and lyophilized.

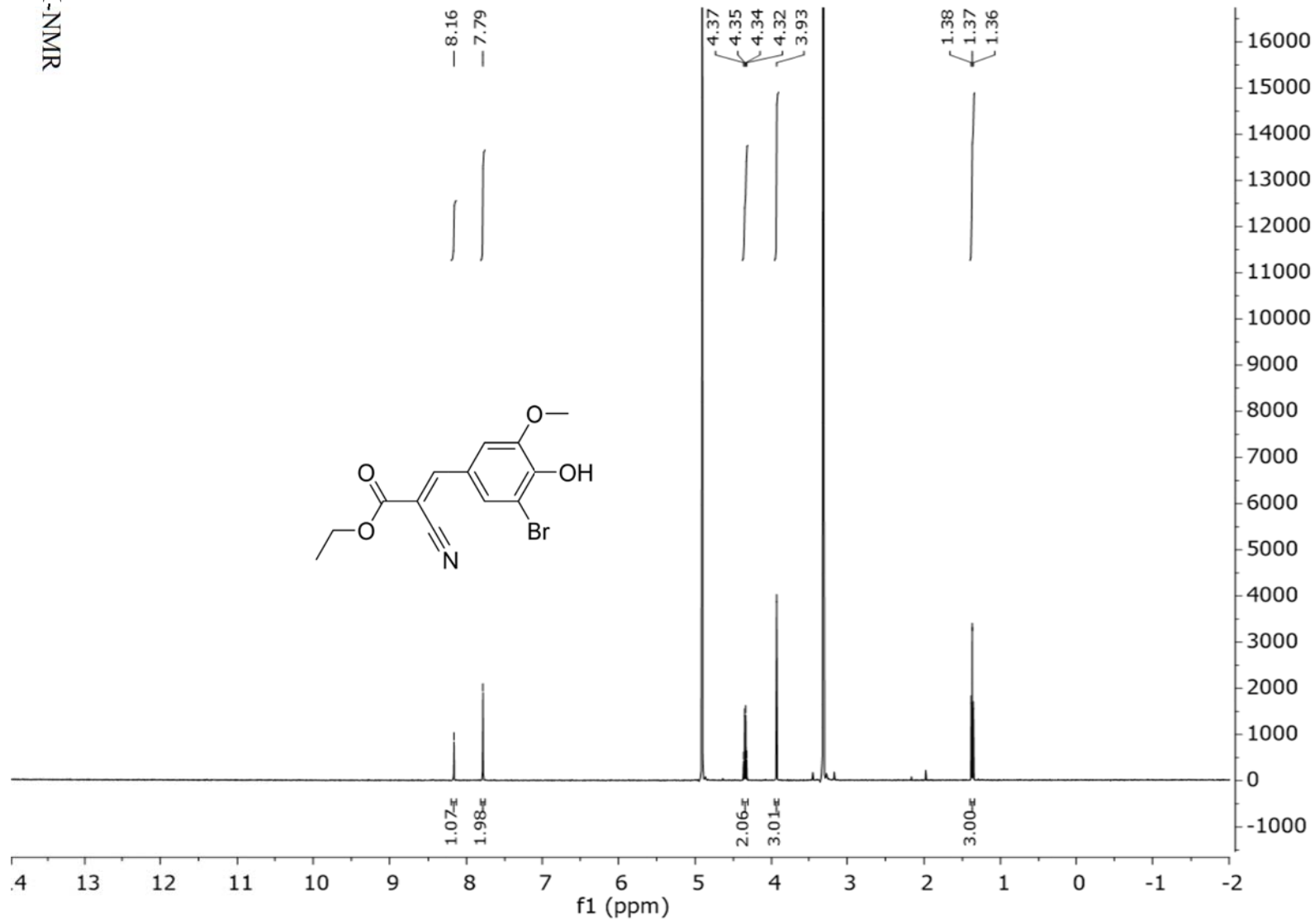
¹H-NMR

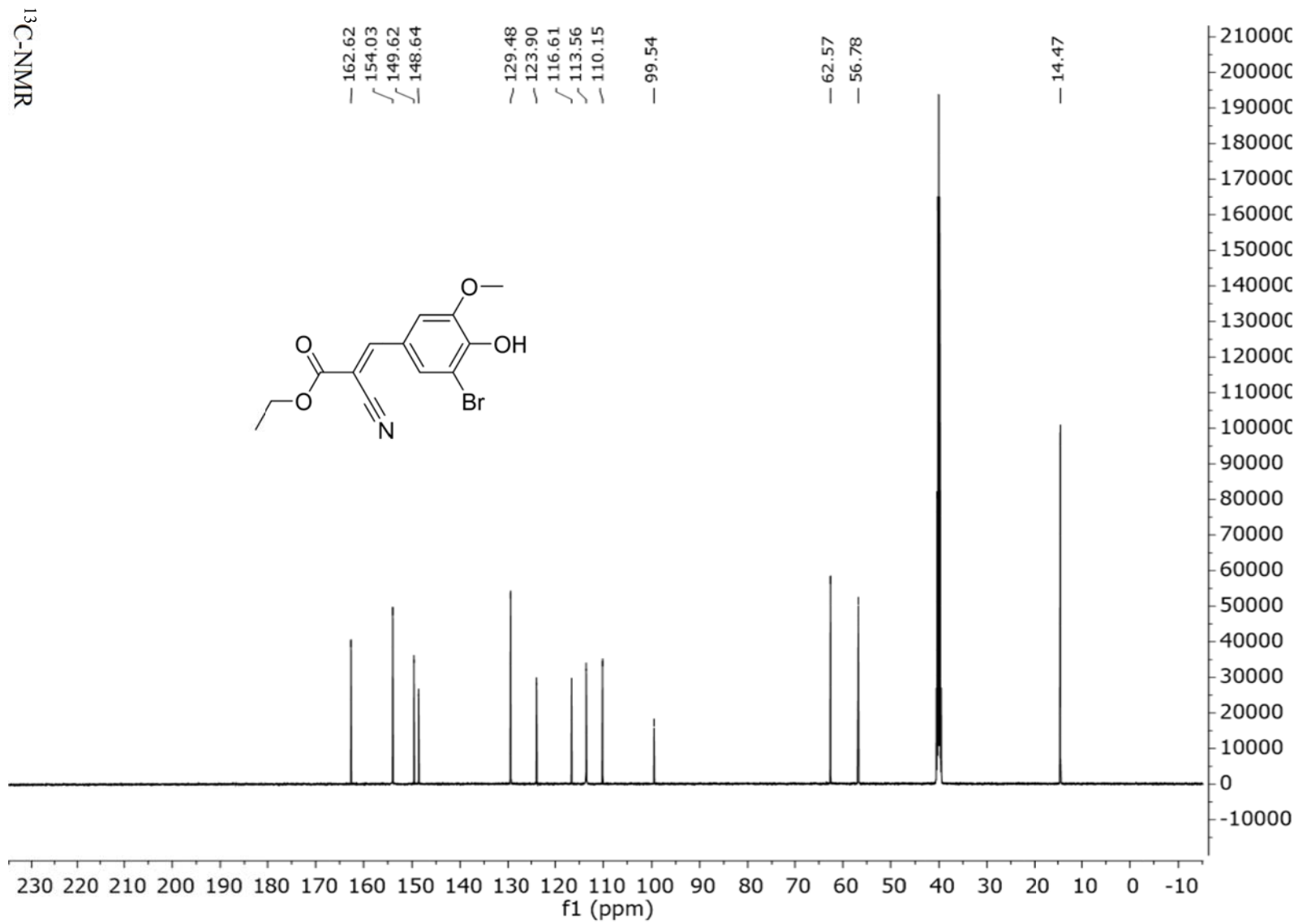


^{13}C -NMR

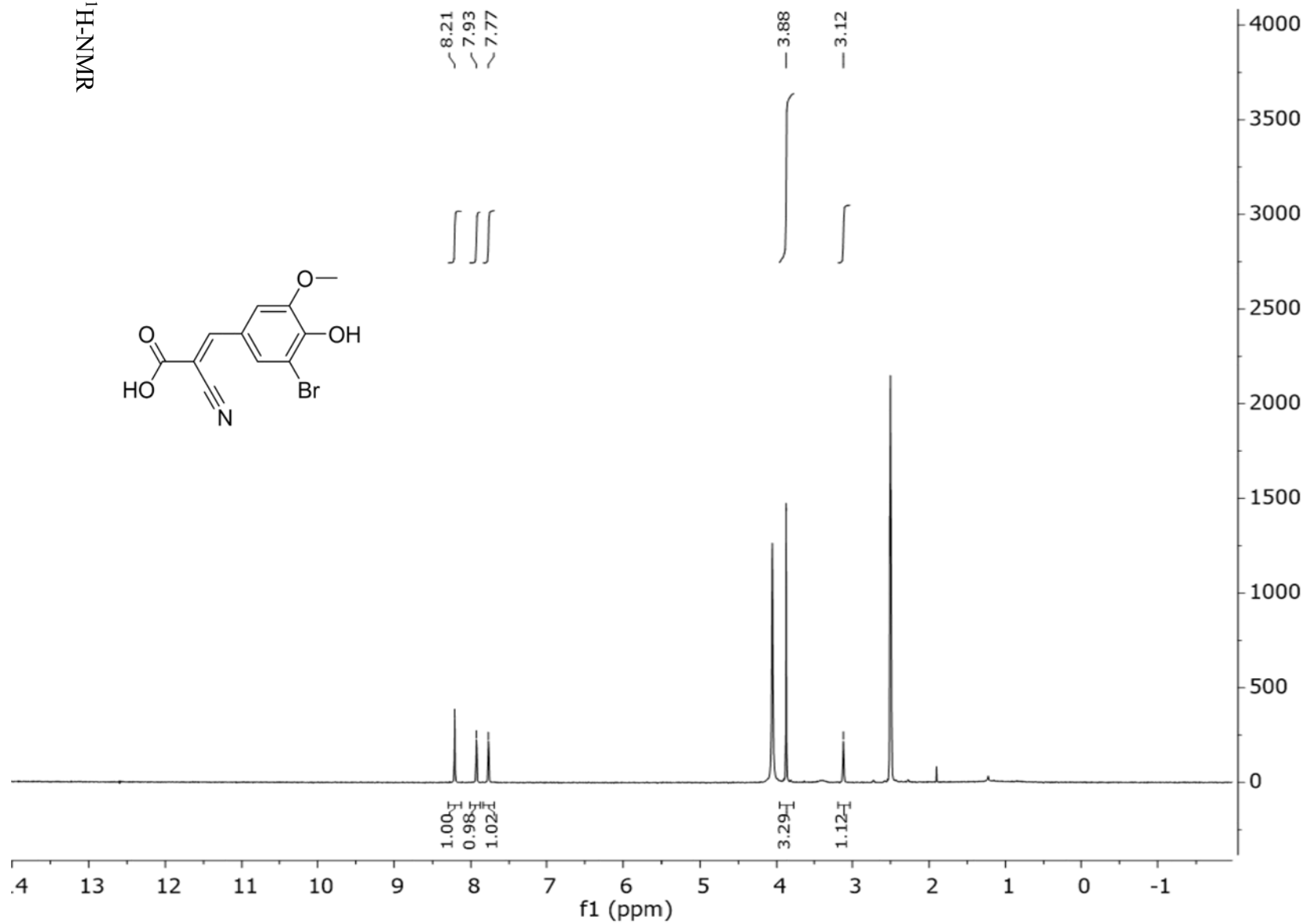
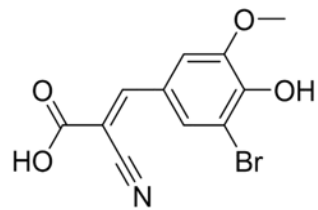


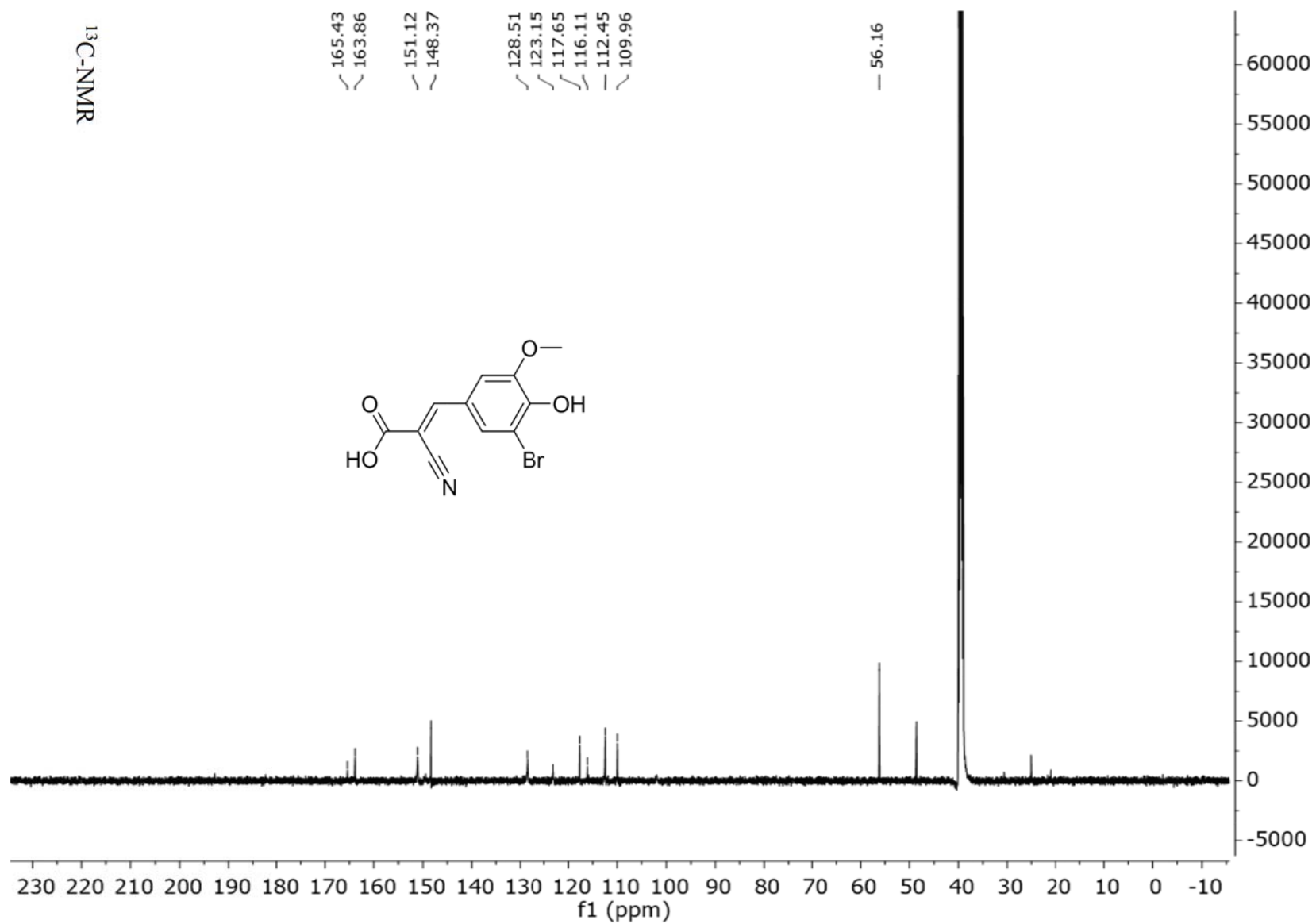
¹H-NMR



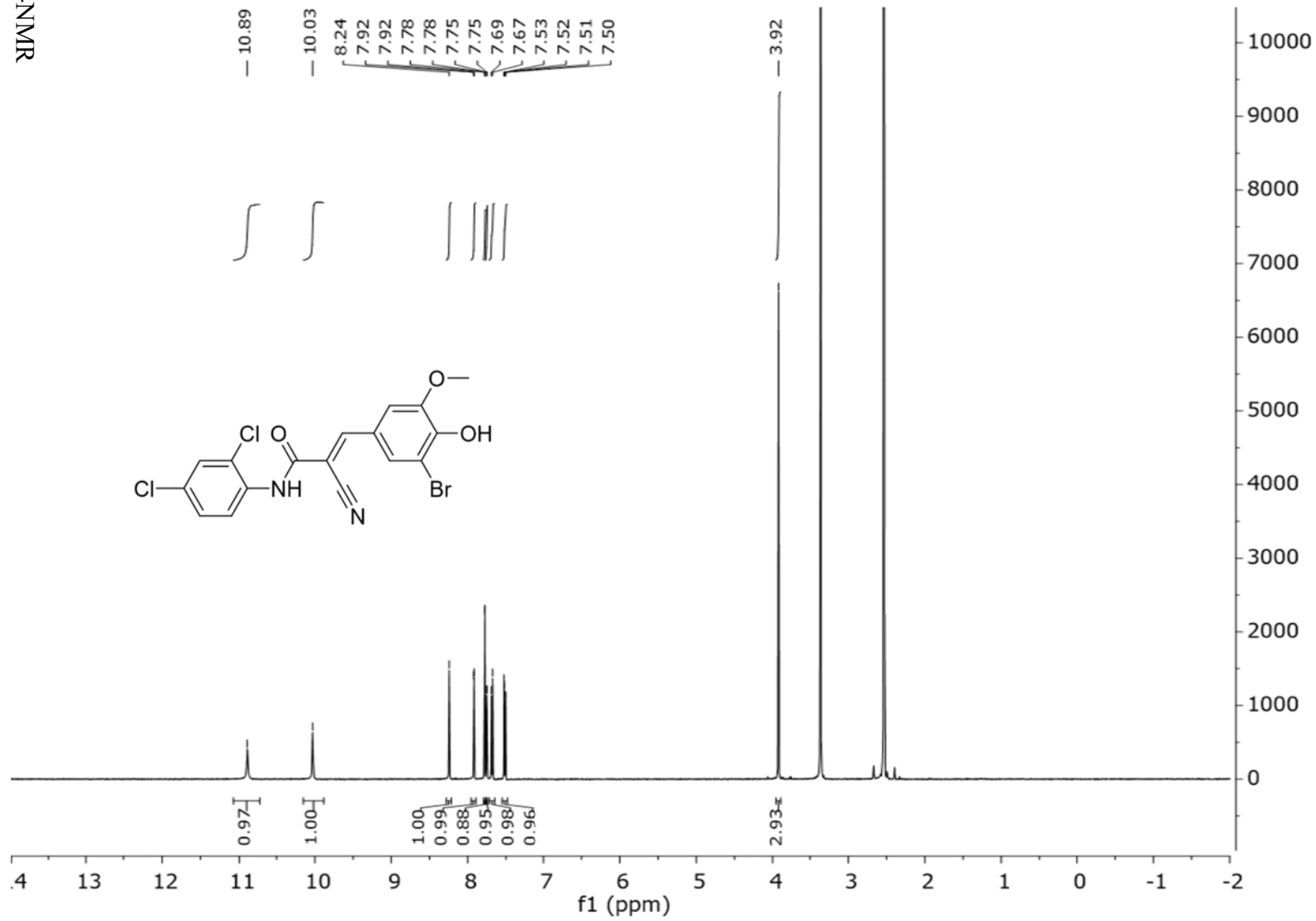


¹H-NMR

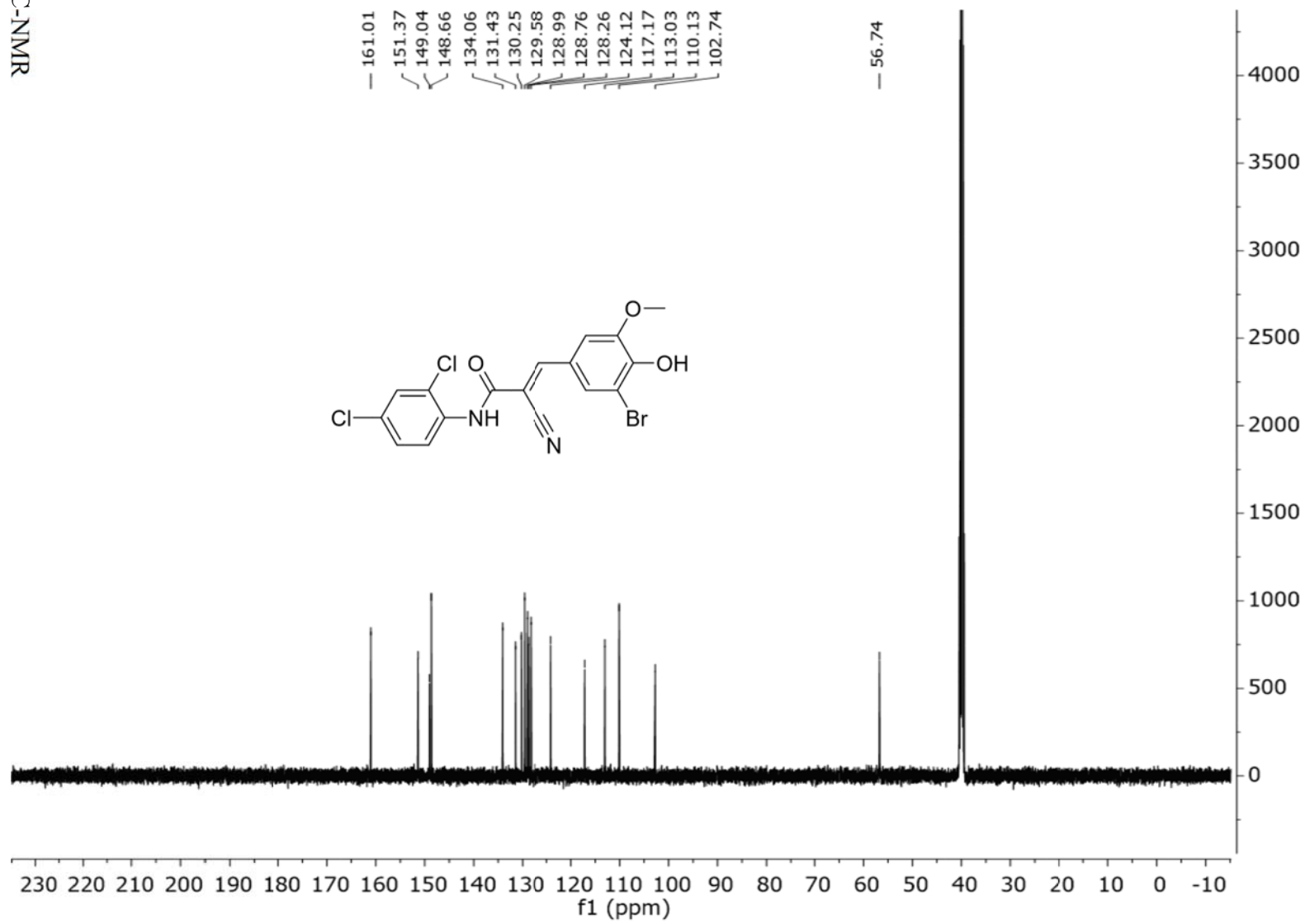


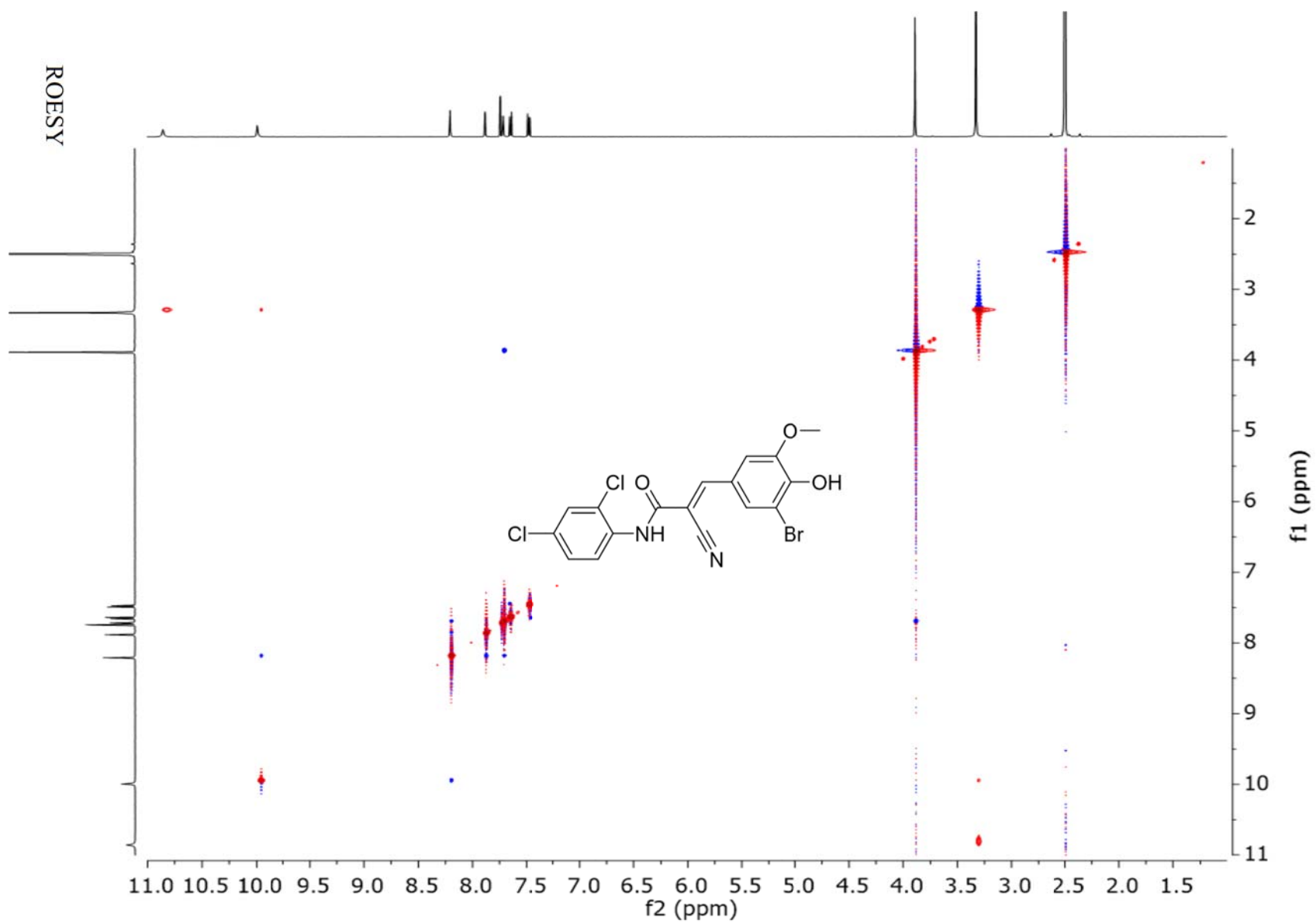


¹H-NMR



^{13}C -NMR





REFERENCES

- (1) Meanwell, N. A. *J. Med. Chem.* **2011**, *54*, 2529.
- (2) Leeson, P. D.; Springthorpe, B. *Nat. Rev. Drug Discov.* **2007**, *6*, 881.
- (3) Roughley, S. D.; Jordan, A. M. *J. Med. Chem.* **2011**, *54*, 3451.
- (4) Chin, P. C.; Liu, L.; Morrison, B. E.; Siddiq, A.; Ratan, R. R.; Bottiglieri, T.; D'Mello, S. R. *J. Neurochem.* **2004**, *90*, 595.
- (5) Zhou, W.; Li, H. B.; Xia, C. N.; Zheng, X. M.; Hu, W. X. *Bioorg. Med. Chem. Lett.* **2009**, *19*, 1861.

Chapter 4

MECHANISTIC AND FUNCTIONAL CHARACTERIZATION OF COMPOUND 37

Chapter 4

MECHANISTIC AND FUNCTIONAL CHARACTERIZATION OF COMPOUND 37

Introduction

An important part of any tool and/or drug discovery process includes assessment of the biological relevance and mechanism of action of the identified compound. This can be accomplished through characterization of specificity, mode of inhibition, and affected biological phenotypes. Here, we adapted the fluorescence assay to characterize specificity against different sulfotransferases, kinetic analysis to understand the mode of inhibition, and various biochemical approaches to study the potential mechanism of action of compound **37**. As proof of principle, we decided to examine the ability of **37** to affect the processes of astrogliosis. We tested the ability of **37** to inhibit the cell-surface expression CS-E on astrocytes, the key cell type involved in generating growth inhibitory CSPGs. Furthermore, we analyzed the disaccharide content of CSPGs secreted from astrocytes and determined that the treatment of astrocytes with **37** attenuated the growth inhibitory effects of the secreted CSPGs. Compound **37** was then characterized *in vitro* and *in vivo* in pharmacokinetic models, including microsomal stability, cytochrome p450 inhibition, and intravenous injection into rats.

Results and Discussion

To determine the specificity of these inhibitors, we utilized the fluorescent, enzyme-coupled screen because it allowed for an adaptable method to compare sulfotransferases with different substrates. When screened in the fluorescence assay, there was a clear dichotomy between the tested sulfotransferases with glycosaminoglycan substrates versus aromatic substrates (Figure 4.1). In this assay, **37** had an $IC_{50} = 2.43 \mu M$ versus Chst15, while against other closely related chondroitin sulfotransferases, Chst11 and Ust, it had IC_{50} 's of 2.45 and 2.02 μM respectively. Versus heparan sulfate 3-O sulfotransferase (Hs3st1), a sulfotransferase for HS GAGs, it also had a similar IC_{50} value of 2.51 μM . However, when **37** was tested against cytosolic sulfotransferases Sult1e1, Sult2b1a, and Sult2b1b, IC_{50} 's increased 7.6- to 17.5-fold. Along with this, the coupled enzyme, Sult1c1, also a cytosolic sulfotransferase, showed no dose response. Due to the similarity of the substrates and the sequence homology amongst GAG sulfotransferases (26-81%),¹⁻¹⁰ the result of **37** also inhibiting other glycosaminoglycan sulfotransferases was not completely unexpected. However, these compounds can still be valuable for the study of the biological roles of GAG sulfation as there are no reported cell-permeable, GAG sulfotransferase inhibitors.

To study the mechanism of inhibition of Chst15 by **37** in greater detail, we utilized the fluorescent assay to examine the type of inhibition against CS-A and PAPS. Analysis of the kinetic data revealed that **37** displayed different modes of inhibition depending on the substrate. When CS-A was the varied substrate, intersecting lines in the second

quadrant of the Lineweaver-Burk plot indicated a mixed mode of inhibition, where both K_m and V_{max} shifted as a result of the inhibitor being present with a $K_i = 1.09 \mu\text{M}$ and K_i'

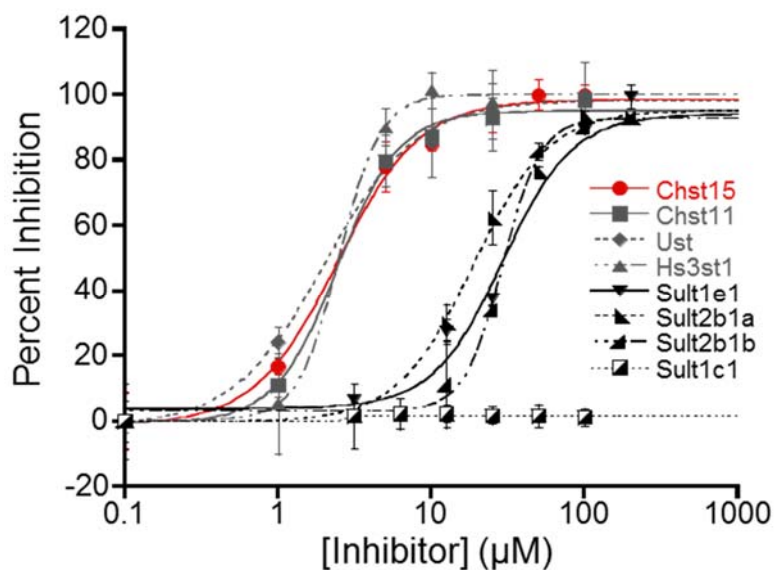


Figure 4.1. Compound **37** selectively inhibits glycosaminoglycan sulfotransferases. Various glycosaminoglycan and cytosolic sulfotransferases were assayed against inhibitor **37** using the fluorescence assay. Of the sulfotransferases tested, **37** was ~12-fold more selective for glycosaminoglycan sulfotransferases compared to cytosolic sulfotransferases. Values reported as the mean \pm S.E.M ($n = 3$).

= $2.14 \mu\text{M}$. In contrast, when PAPS was the varied substrate, lines intersecting on the y-axis signified a competitive mode of inhibition with a $K_i = 2.58 \mu\text{M}$ (Figure 4.2 and Table 4.1). The inhibitor's competitive nature against PAPS offers an explanation for its cross reactivity with other sulfotransferases.

We hypothesized that the α,β -unsaturated acrylamide moiety could serve as an electrophilic site for a nucleophilic amino acid side chain. Indeed, compounds with similar cyanoacrylamide structures have been reported to inhibit noncatalytic cysteines on kinases

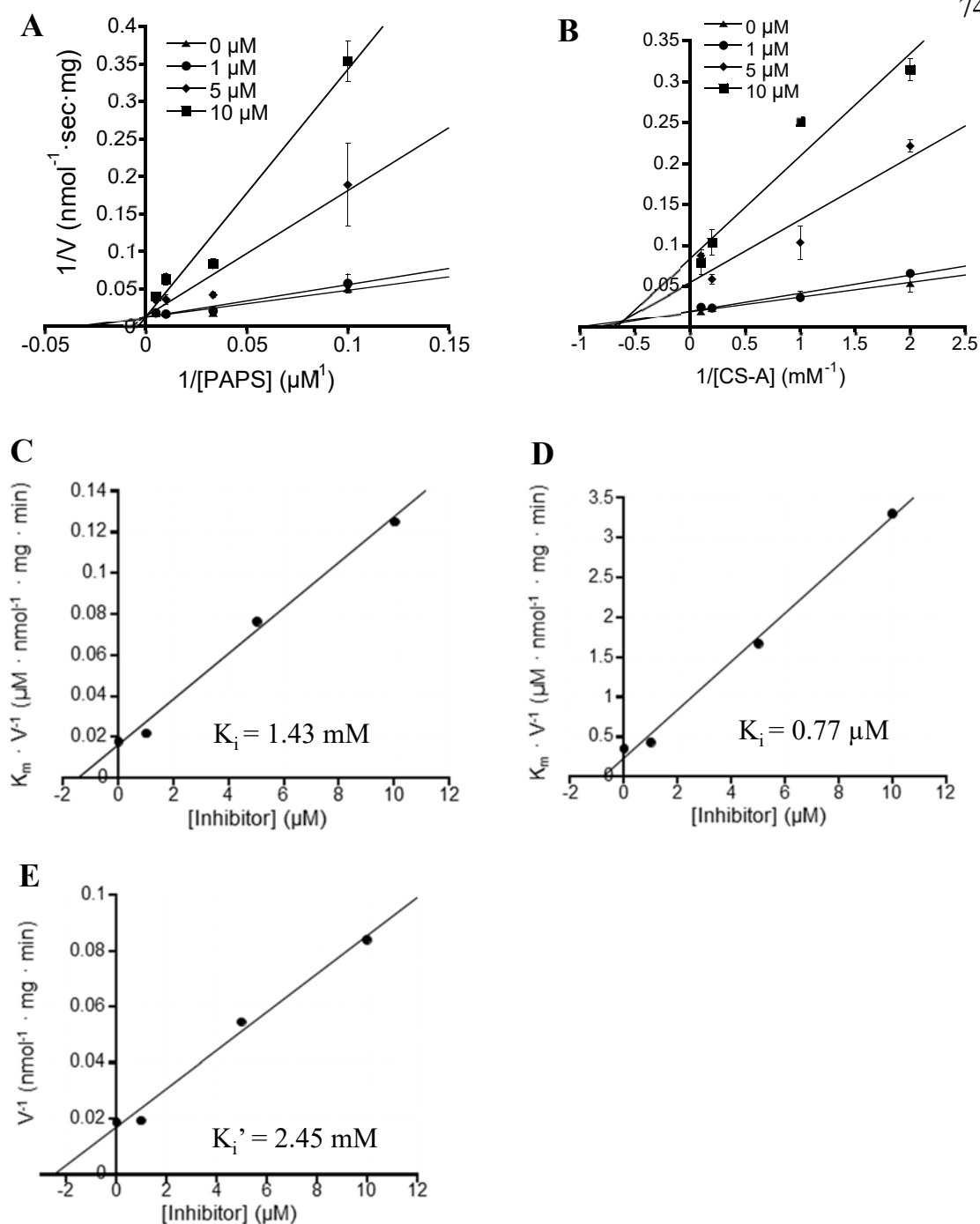


Figure 4.2. Steady-state kinetic analysis reveals competitive and mixed modes of Chst15 inhibition by **37**. Initial reaction velocities were determined by the fluorescence assay with varying substrate concentrations, (A) CS-A and (B) PAPS, at different concentrations of **37**. Double reciprocal plots (Lineweaver-Burk) showed **37** is a competitive inhibitor of PAPS and a mixed inhibitor of CS-A. Values represent the mean \pm S.E.M. ($n = 3$). K_i and K_i' parameters were derived from a secondary plot of K_m/V versus [**37**] for (C) CS-A and (D) PAPS. (E) K_i' was derived from a secondary plot of $1/V$ versus [**37**] for CS-A.

| Varied Substrate | 37 (μM) | k_{cat} (sec^{-1}) | $K_{\text{m,app}}$ (μM) | K_i (μM) | K_i' (μM) |
|------------------|----------------------|--|--------------------------------------|-------------------------|--------------------------|
| PAPS | 0 | 0.126 | 29.49 | 0.77 | - |
| | 1 | 0.125 | 35.86 | | |
| | 5 | 0.114 | 124.57 | | |
| | 10 | 0.121 | 263.12 | | |
| Varied Substrate | 37 (μM) | k_{cat} (sec^{-1}) | $K_{\text{m,app}}$ (mM) | K_i (mM) | K_i' (mM) |
| CS-A | 0 | 0.081 | 0.96 | 1.43 | 2.45 |
| | 1 | 0.078 | 1.12 | | |
| | 5 | 0.028 | 1.40 | | |
| | 10 | 0.018 | 1.49 | | |

Table 4.1. Kinetic parameters of inhibition for Chst15. Kinetic parameters were determined by measuring the initial reaction velocities at varying substrate concentrations in the presence of **37** at concentrations of 0, 1, 5, and 10 μM . k_{cat} and K_{m} were determined from Lineweaver-Burk plots. K_i and K_i' were derived from secondary plots of K_{m}/V versus **37** or $1/V$ versus **37**, respectively.

via a “reversible covalent” mechanism.^{11,12} Consistent with this mechanism, we observed a reduction in the prominent UV-visible absorption band of the cyanoacrylamide **37** ($\lambda_{\text{max}} = 378 \text{ nm}$) when the compound was incubated with a surrogate nucleophile, β -mercaptoethanol (βME), suggesting the formation of covalent adduct **56** (Figure 4.3A and 4.3B). Formation of this adduct was rapidly reversed when **37** was diluted 10-fold into phosphate buffered saline (PBS) without βME (Figure 4.3C). Further evidence for a reversible thioether adduct was obtained by ^1H NMR spectroscopy (Figure 4.3D). Appearance of an aliphatic β -proton in adduct **56** was witnessed upon incubation of **37** with βME . The integrated peak ratio between this proton and the olefin proton in **37** indicated a substrate to adduct ratio of 43:57. Two-fold dilution into PBS shifted this ratio to 63:37, indicating reversion of this equilibrium. In enzymatic activity assays, longer pre-

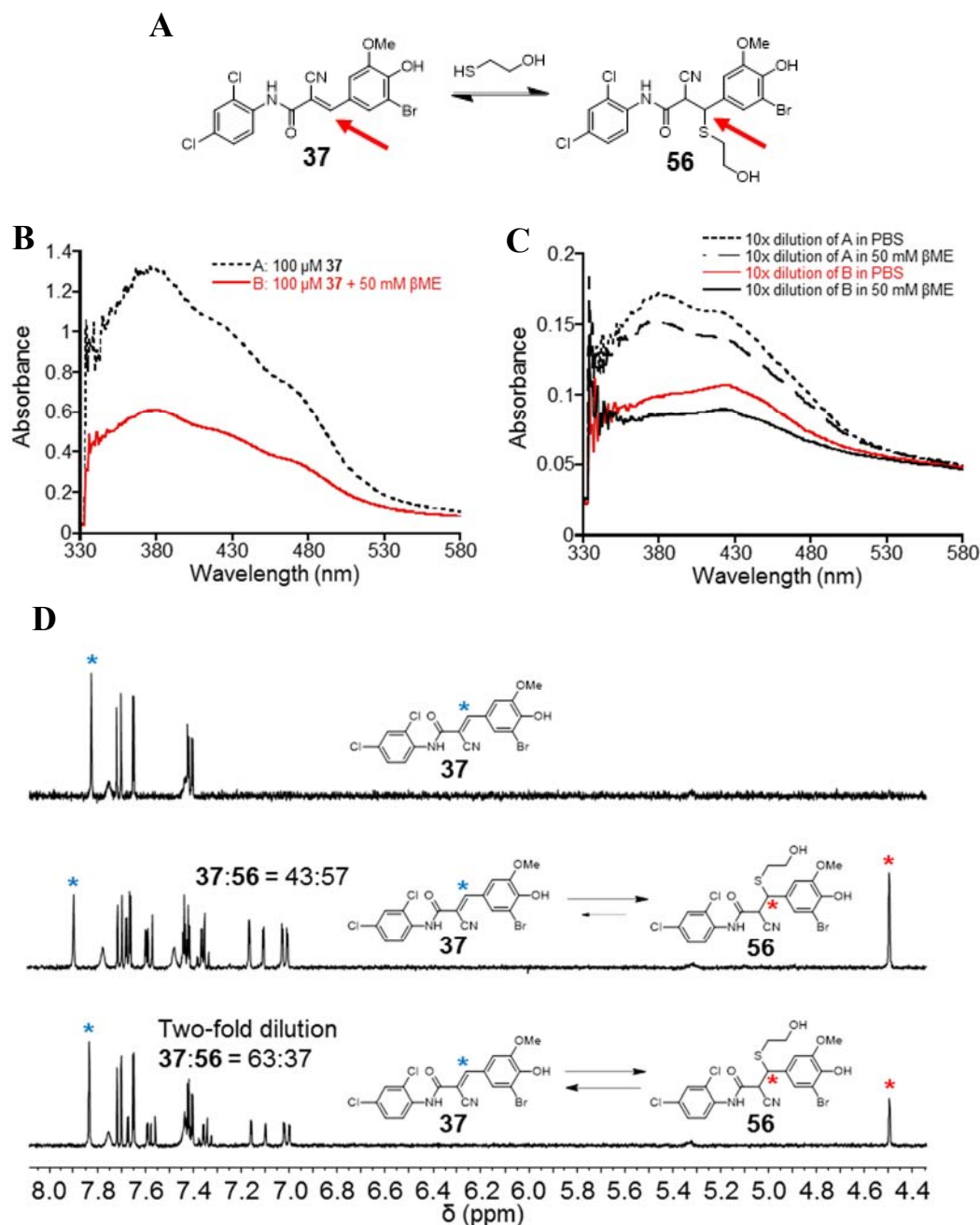


Figure 4.3. Compound **37** exhibits reversible-covalent inhibitor properties when using β -mecaptoethanol (β ME) as a model nucleophile. (A) Conjugate addition reaction of β ME with **37**. (B) UV-Vis absorption ($\lambda_{\text{max}} = 378$ nm) was lost upon incubation of **37** with β ME suggesting conjugation. (C) However, 10-fold dilution of B with PBS led to a reversal in the equilibrium when compared to the dilution of B with β ME. (D) ^1H -NMR showing the appearance of a newly formed aliphatic proton (red) after incubation with β ME indicated conjugation. Upon dilution, the ratio of the integrated olefin (blue) and aliphatic proton signaled a shift in the equilibrium back to **37**.

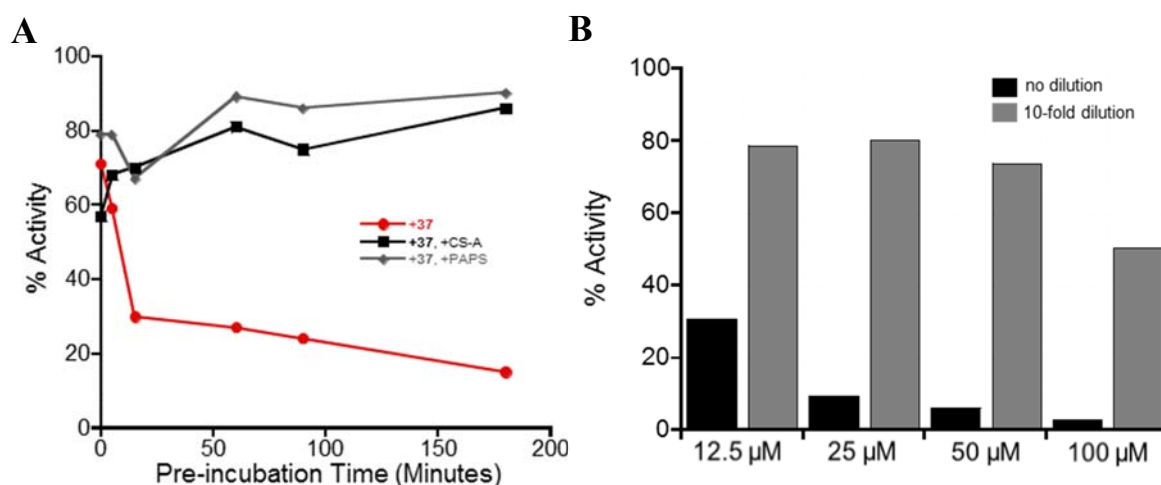


Figure 4.4. **37** inhibitory activity depends on pre-incubation time and shows reversibility. (A) Longer pre-incubation times result in greater inhibition by **37**, supporting irreversible properties. The compound is out competed by the substrates CS-A (1 mM) and PAPS (30 μM). (B) **37** inhibitory activity is not dilution resistant, supporting its reversibility. Chst15 is pre-incubated with **37** at the indicated concentrations (12.5 μM, 25 μM, 50 μM, 100 μM). However, a loss of inhibitory activity is seen in samples diluted 10-fold. This is only expected for reversible compounds where inhibitory activity is only dependent on the final concentration.

incubation times with the inhibitor led to a higher observed inhibition, supporting its covalent nature (Figure 4.4A). Furthermore, greater enzyme activity was restored with 10-fold dilution, indicating a dependence only on final concentration and thus supporting its reversibility (Figure 4.4B).

The ability to inhibit GAG sulfation after CNS injury could have exciting implications for neural regeneration (Figure 4.5). Studies have established that the CS-E sulfation motif is upregulated on CSPGs after injury and inhibits the growth of sensory axons.^{13,14} To explore the effects of these small molecules in a cellular context, we decided to examine if **37** treatment could lead to the hypothesized decrease in CS sulfation. To this

end, we first treated NIH3T3 fibroblasts and Neu7 astrocytes with **37** and immunofluorescently stained with a monoclonal CS-E antibody developed previously.¹³ Quantification of the images revealed a significant, dose-dependent decrease in CS-E

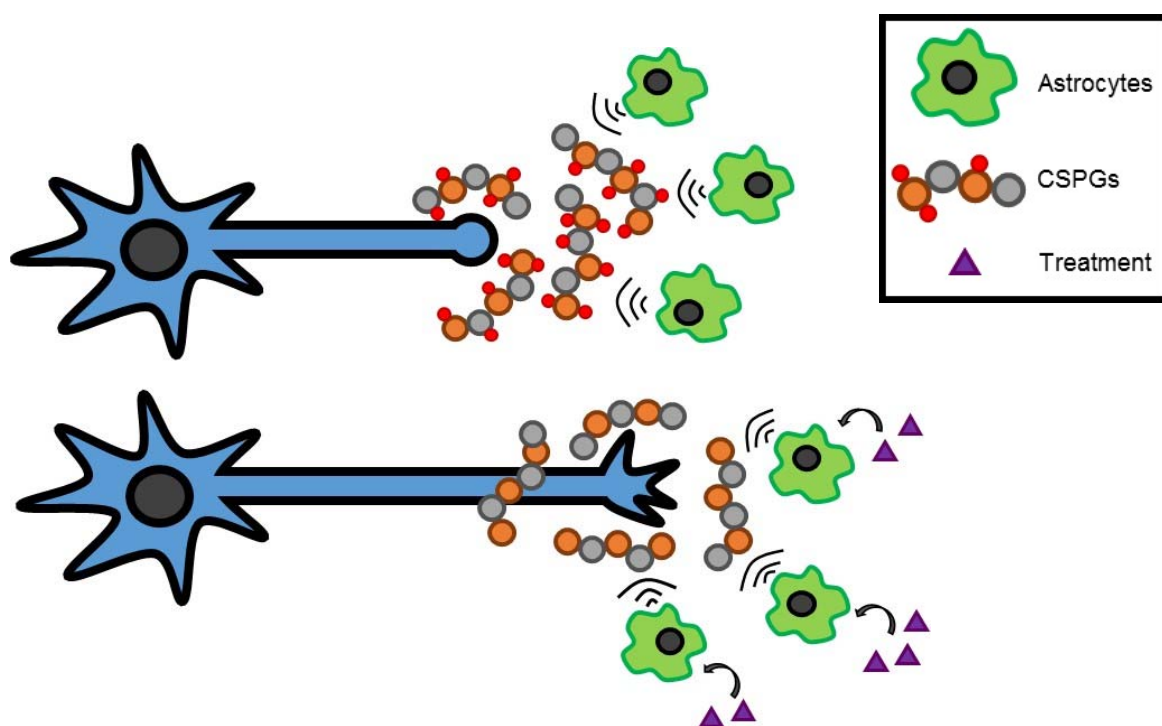


Figure 4.5. Representation of astroglial inhibition of axonal growth after neuronal injury and the potential therapeutic mechanism of small molecule treatment. Reactive astrocytes release over-sulfated CSPGs that inhibits axonal growth in response to neuronal insult. Small-molecule treatment at the site of injury may reduce the sulfation of secreted CSPGs and potentiate axonal regrowth.

staining when treating both cell types with **37** as compared to untreated cells (Figure 4.6).

The effect was maximal at 25 μ M of **37**. Likewise, unsurprisingly, treatment also decreased HS expression seen when staining with a monoclonal HS antibody (Figure 4.7). These decreases in staining were also recapitulated by treatment of both cell types with sodium chlorate, a PAPS synthetase inhibitor often used as a universal sulfation inhibitor.

However, its global effects on sulfation, oxidative properties, and use in millimolar concentrations precludes its therapeutic application, again highlighting the significance of these sulfotransferase inhibitors.^{15,16}

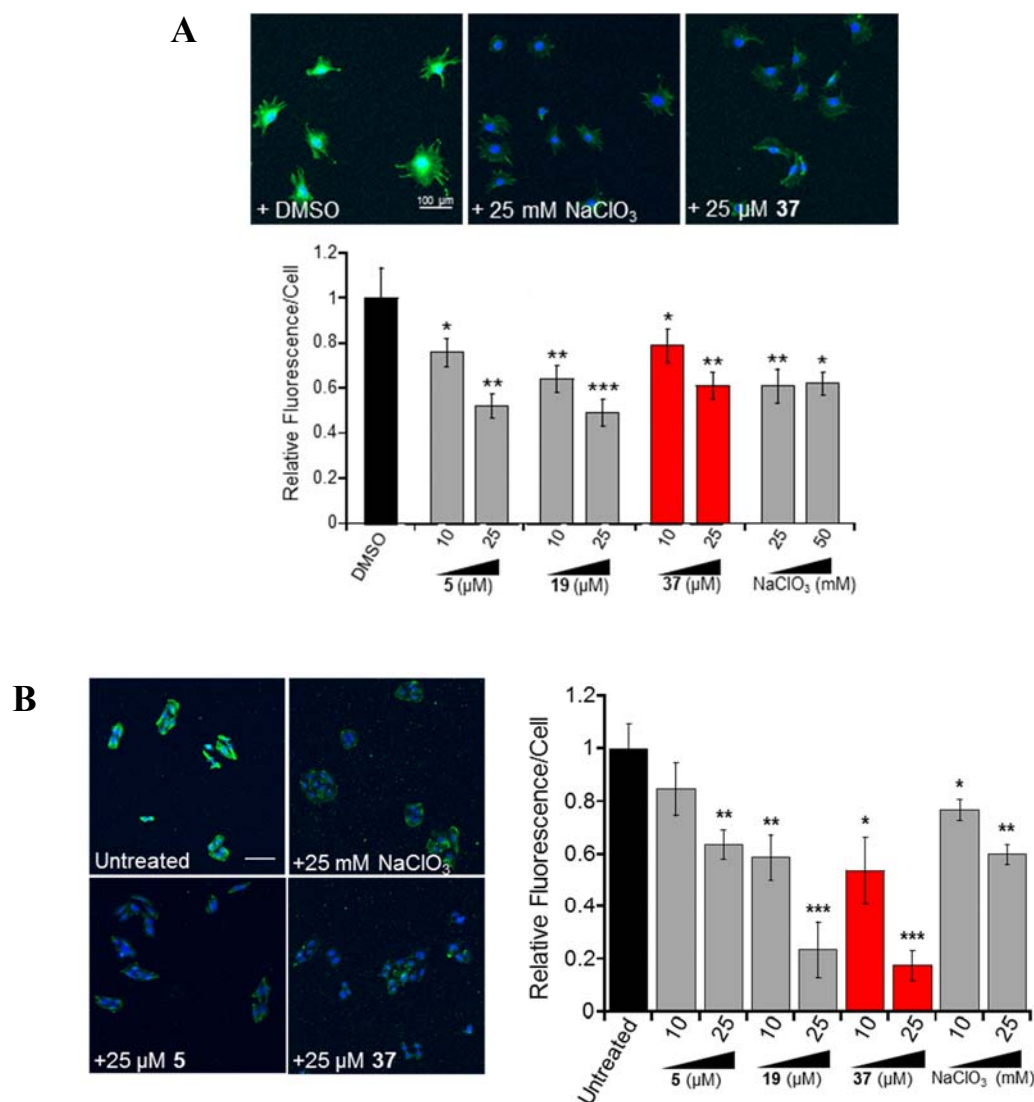


Figure 4.6. Treatment of NIH/3T3 fibroblasts and Neu7 astrocytes with **37** leads to a significant, dose-dependent decrease in CS-E expression. Cells were grown in the presence of inhibitors for 40 hours and stained with a monoclonal CS-E antibody (green) and DAPI (blue, astrocytes only). Cells were imaged and quantified by fluorescence microscopy (scale bar = 100 μ M). Treatment of inhibitors were compared with untreated cells (One-way ANOVA, * P < 0.05, ** P < 0.01, *** P < 0.001). Quantification from three experiments is shown (n = 20-100 cells per experiment).

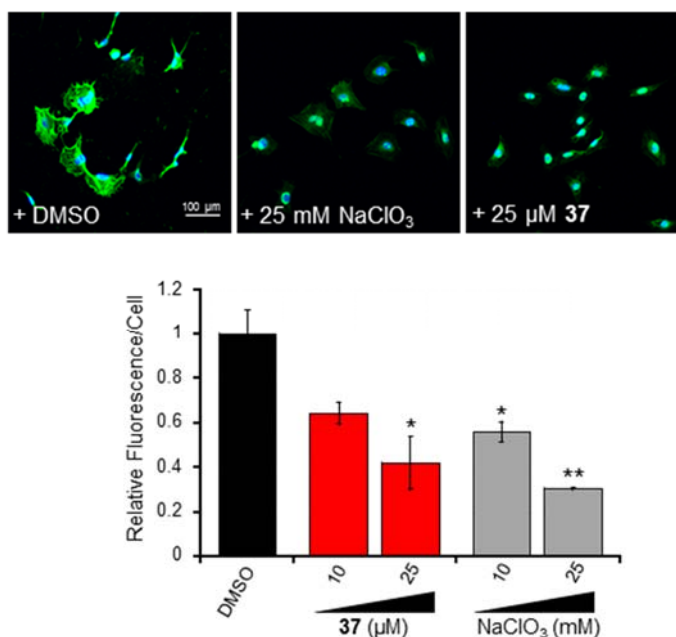


Figure 4.7. Treatment of NIH/3T3 fibroblasts with **37** leads to a significant, dose-dependent decrease in HS expression. NIH/3T3 fibroblasts were grown in the presence of inhibitors for 40 hours and stained with a polyclonal HS antibody (green) and DAPI (blue). Cells were imaged and quantified by fluorescence microscopy (scale bar = 100 μM). Treatment of inhibitors were compared with untreated cells (One-way ANOVA, * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$). Quantification from three experiments is shown ($n = 20$ -100 cells per experiment).

After spinal cord injury, reactive astrocytes enter the site of injury and secrete CSPGs as a means to sequester further damage. The Neu7 astrocytic cell line has been previously characterized to secrete neuronal growth inhibiting CSPGs.^{17,18} To characterize the secreted CSPGs, we purified them from Neu7 conditioned media, digested them with chondroitinase ABC (ChABC), and then analyzed their disaccharide content by reversed-phase ion-pair high-performance liquid chromatography (RPIP-HPLC).¹⁹ CSPGs collected from astrocytes treated with **37** showed a clear decrease in overall sulfation. Not unexpectedly, each CS sulfation motif was affected to varying degrees, with the CS-E disaccharide seeing the largest decrease of $67.8\% \pm 11.0\%$ at 25 μM of **37** as compared to

CSPGs from untreated cells (Figure 4.8 and Table 4.2). These studies indicated that the discovered inhibitors decreased sulfation on both surface-expressed and secreted CSPGs.

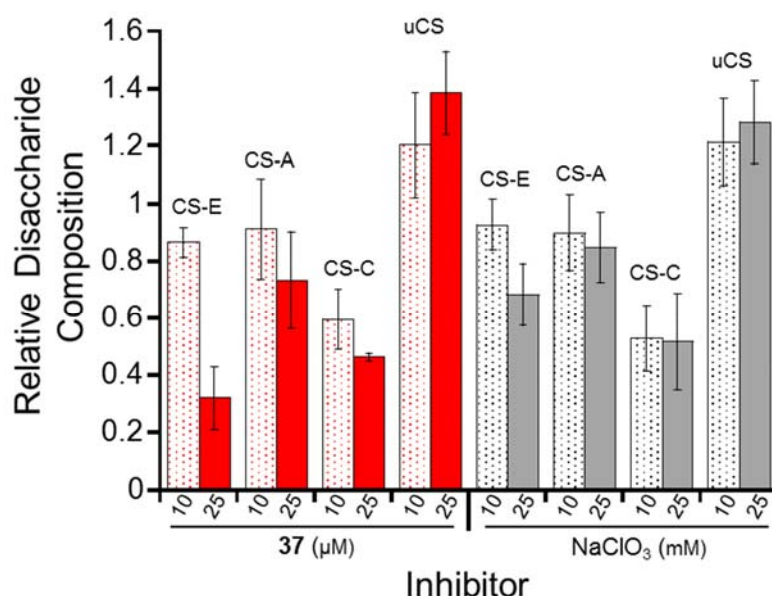


Figure 4.8. CSPGs from Neu7 astrocytes treated with inhibitor **37** show decreased levels of CS-E and other sulfation motifs. CS disaccharide composition was analyzed by reverse phase ion-pairing high pressure liquid chromatography (RPIP-HPLC). CS-D was not detected. uCS, unsulfated CS. Values are normalized to the disaccharide content on CSPGs from untreated astrocytes and are reported as the mean \pm S.E.M. ($n = 3$).

| Disaccharide | Distribution (%) in Neu7 Secreted CSPGs with | | | | |
|--------------|--|------------------|------------------|--------------------------|--------------------------|
| | No Treatment | 10 μ M 37 | 25 μ M 37 | 10 mM NaClO ₃ | 25 mM NaClO ₃ |
| uCS | 38.77 \pm 4.19 | 48.74 \pm 7.35 | 63.18 \pm 6.59 | 49.40 \pm 6.14 | 54.18 \pm 6.05 |
| CS-A | 37.05 \pm 3.07 | 33.82 \pm 6.44 | 27.20 \pm 6.25 | 33.34 \pm 4.91 | 31.42 \pm 4.50 |
| CS-C | 12.97 \pm 3.10 | 7.93 \pm 1.36 | 7.74 \pm 0.18 | 6.87 \pm 1.48 | 6.73 \pm 2.19 |
| CS-D | n.d. | n.d. | n.d. | n.d. | n.d. |
| CS-E | 11.20 \pm 0.65 | 9.70 \pm 0.59 | 3.61 \pm 1.22 | 10.40 \pm 1.00 | 7.67 \pm 1.22 |

n.d. = not detected

Table 4.2. Quantification of the percent disaccharide composition of CSPGs secreted from untreated and treated astrocytes. CS disaccharide composition was analyzed by reverse phase ion-pairing high pressure liquid chromatography (RPIP-HPLC). CS-D was not detected. uCS, unsulfated CS. Values are reported as the mean \pm S.E.M. ($n = 3$).

As a potential therapeutic agent, treatment of the site of injury would generate an environment in which all cell types (neurons, astrocytes, oligodendrocytes, microglia) would encounter the compound.²⁰ While the main target for the inhibitors are the reactive astrocytes secreting CSPGs, it was important to determine the effects on neuronal growth directly. When treated with inhibitors directly, P5 rat dorsal root ganglion neurons (DRGs) did not show any significant changes in neurite outgrowth (Figure 4.9). Interestingly, compound **5** potentially shows slight growth promoting effects, which agrees with its previously reported neuroprotective activity by inhibiting cRaf.²¹

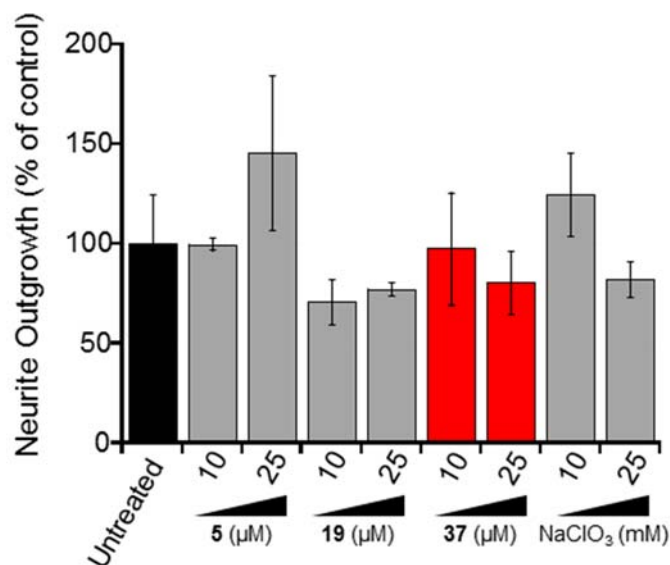


Figure 4.9. DRG treatment with **37** leads to no significant effect in neurite outgrowth. P5 rat DRGs were grown in the presence of inhibitors for 24 hours and imaged by fluorescence microscopy. DRG outgrowth with inhibitor treatment was compared with untreated cells (One-way ANOVA). Quantification from three experiments is shown ($n = 20-80$ cells per experiment).

As an *in vitro* model of the spinal cord environment post-injury, DRGs were grown on a substratum coated with the CSPGs purified from Neu7 astrocytes. To our delight, compared to DRGs grown on CSPGs from untreated Neu7 astrocytes, there was a significant, dose-dependent promotion of neurite outgrowth of up to 30.7% of DRGs grown on CSPGs from **37** treated astrocytes (Figure 4.10). In fact, treatment with 25 μ M **37** reverses the inhibitory effects produced by these CSPGs. This dose-dependent trend in outgrowth promotion was also seen with sodium chlorate treatment, highlighting again that sulfation plays key role in CSPG-mediated axonal inhibition.

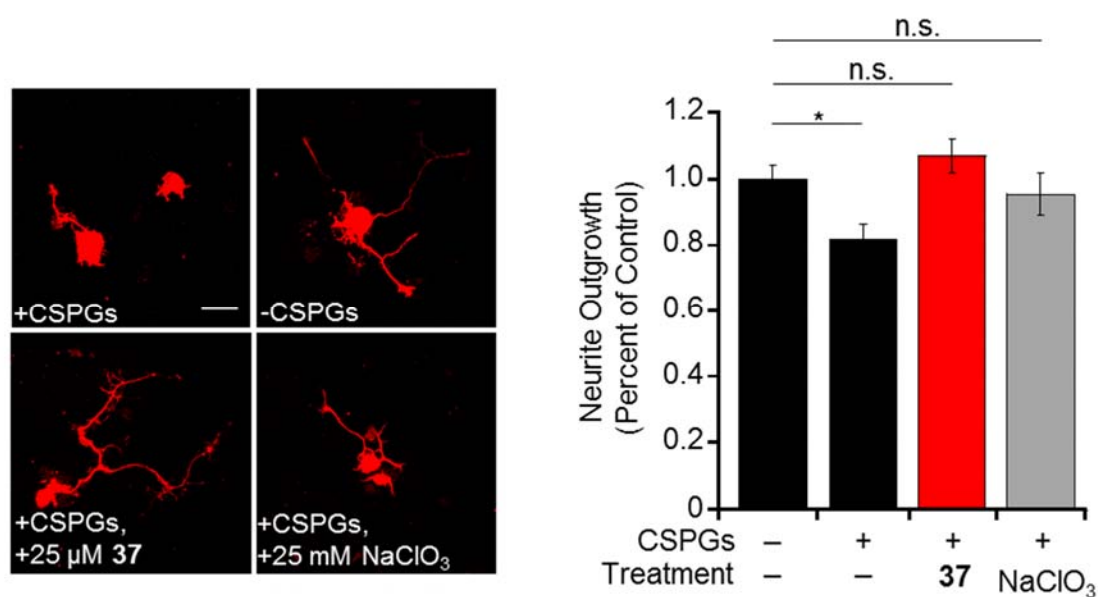


Figure 4.10. The inhibitory effects of CSPGs secreted from Neu7 astrocytes are attenuated by treatment of the cells with **37** or the general sulfation inhibitor sodium chlorate. P5 rat DRGs were grown on substrata coated with or without CSPGs isolated from astrocytes following treatment as indicated. Cells were imaged by fluorescence microscopy (scale bar = 50 μ m), and neurite outgrowth quantified. Data were normalized relative to DRGs grown on CSPGs from untreated astrocytes. One-way ANOVA, * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$. Quantification from three experiments is shown ($n = 20$ -80 neurons per experiment).

Similarly, in another *in vitro* model of the glial scar environment, we examined if the inhibitor treatment of astrocytes could reduce the repulsion of axons seen when encountering a boundary created by secreted CSPGs in a boundary assay format. As hypothesized, CSPGs from untreated astrocytes formed an inhibitory zone that was highly repellent to DRG axons (Figure 4.11). In contrast, axons more freely crossed into boundaries with CSPGs treated with **37** as there was no significant difference in axon crossing compared to a boundary containing no CSPGs.

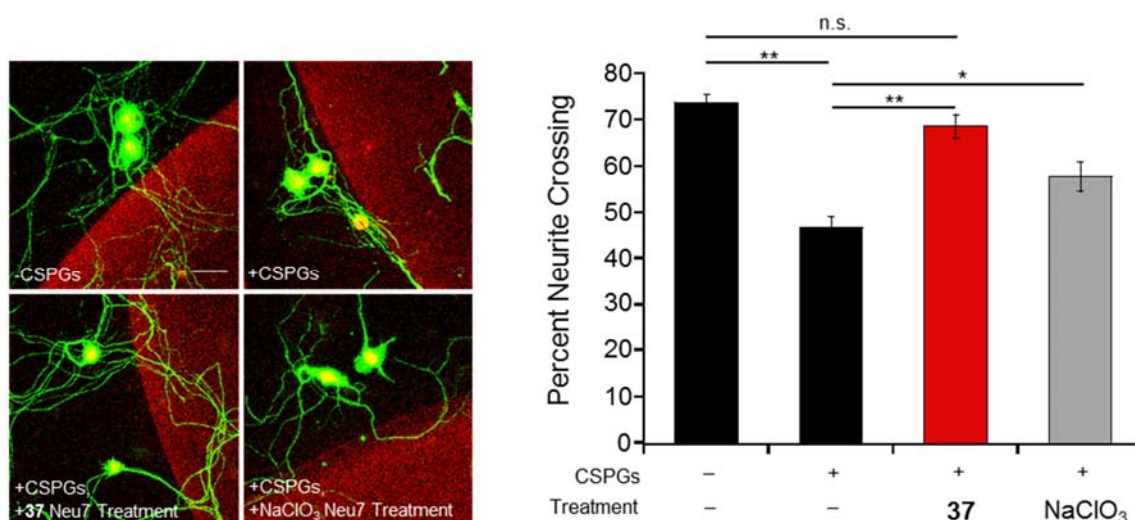


Figure 4.11. CSPGs from astrocytes treated with **37** repel fewer neurites compared to CSPGs from untreated astrocytes. A boundary assay was performed in which purified CSPGs from Neu7 astrocytes (5 $\mu\text{g/mL}$) or PBS control were mixed with Texas Red and spotted on poly-lysine coated coverslips. P5 rat DRGs were grown on the coverslips and quantified for their ability to cross the CSPG-containing boundary (scale bar = 50 μm). CSPGs from inhibitor treated astrocytes were significantly less inhibitory toward DRGs (25 μM **37**, 25 mM NaClO₃). One-way ANOVA, n.s. = not significant, * $P < 0.05$, ** $P < 0.0001$. Quantification from five experiments is shown ($n = 20$ -40 neurons per experiment).

As a prelude to conducting *in vivo* injury experiments, we tested **19** and **37** in *in vitro* liver microsomal stability and cytochrome p450 inhibition assays to profile absorption, distribution, metabolism, and elimination (ADME) properties. The compound showed good stability in human, mouse, and rat liver microsomes based on intrinsic clearance (Table 4.3). Encouragingly, of the four cytochrome p450 isoforms tested, three were not appreciably inhibited at compound **19** concentrations as high as 25 μM (Table 4.4). The pharmacokinetic (PK) properties of **37** were also evaluated directly in rats. After intravenous injection of compound **37** at a dose of 3.0 mg/kg, the compound had a moderate clearance of 20.61 mL/min/kg, moderate volume of distribution of 0.97 L/kg, and short terminal half-life of 1.58 h (Table 4.5). This data seemed to indicate that using a direct, as opposed to intravenous, application of the drug may be more efficacious for future *in vivo* injury studies.

Table 4.3. *In Vitro* Microsomal Stability of **19** and **37**

| Microsome Species | CL _{int} ($\mu\text{L}/\text{min}/\text{mg}$) | |
|-------------------|--|-----------|
| | 19 | 37 |
| Human | <6.9 | <6.9 |
| Mouse | <6.9 | <6.9 |
| Rat | <6.9 | <6.9 |

Table 4.4. *In Vitro* Cytochrome P450 Inhibition by **19**

| CYP Isoform | IC ₅₀ (μM) |
|-------------|------------------------------------|
| CYP2C9 | 0.7 \pm 0.1 |
| CYP2D6 | > 25 |
| CYP3A4 II | > 25 |
| CYP3A4 I | > 25 |

Table 4.5. Pharmacokinetic Parameters of **37** in Rats

| Route | Dose (mg/kg) | AUC _{0-∞} (h·nM) | CL (mL/min/kg) | V _{ss} (L/kg) | C _{max} (nM) | T _{max} (h) | T _{1/2} (h) |
|-------|-----------------|------------------------------|-------------------|---------------------------|--------------------------|-------------------------|-------------------------|
| IV | 3.0 | 5486 | 21 | 0.97 | 11847 | 0.03 | 1.6 |

* IV = intravenous injection, AUC = area under the curve, CL = plasma clearance, V_{ss} = volume of distribution, C_{max} = maximum plasma concentration, T_{max} = time of maximum plasma concentration, T_{1/2} = half life

Conclusion

In conclusion, here we have described the *in vitro* and *in vivo* characterization of the most potent, optimized compound, **37**. We showed that compound **37** did not only inhibit Chst15, but also other similarly related glycosaminoglycan sulfotransferases involved in the biosynthesis of HS and CS. Kinetic analysis showed that the compound inhibited PAPS in a competitive fashion, which offers an explanation for the nonspecificity seen against other GAG sulfotransferases. Also, it was shown that compound **37** likely inhibits through a unique “reversible-covalent” mechanism based on its reactivity with a model nucleophile and time-dependent inhibition. Excitingly, when used in a cellular context, the inhibitor was able to reduce the sulfation of cell-surface expressed and secreted GAGs. The ability of these sulfated GAGs to potentiate neuronal inhibition was significantly decreased when they were secreted from inhibitor treated astrocytes. With its proven ability to inhibit GAG sulfation and in preparation of further *in vivo* collaborations, we determined that the PK parameters were good *in vitro*, while *in vivo* tests suggested direct application of the inhibitor may be most efficacious. Ongoing *in vivo* collaborations

to test the molecule's biological relevance in neuronal injury models will be an important next step in the development of this tool into a potential therapeutic.

Experimental Methods

Expression and Purification of Sult2b1a and Sult2b1b. BL21(DE3) (Novagen) harboring the pGEX-6P-3-Sult2b1a or pGEX-6P-3-Sult2b1b constructs were kindly provided by Dr. Lars Pedersen (NIEHS/NIH). Cells were grown in LB medium (1L) at 37°C. When cells reached an OD₆₀₀ of 0.8, cells were moved to 16°C for 30 minutes. Isopropyl β -D-1-thiogalactopyranoside (IPTG) was added to the cells at the final concentrations of 0.2 mM after the 30 minutes and cells were incubated for an additional 18 hours at 16°C. The pelleted cells were lysed in cold PBS (10 mM Na₂HPO₄, 1.8 mM KH₂PO₄ pH 7.3, 140 mM NaCl, 2.7 mM KCl, and 1x Complete[™] protease inhibitors (Roche) by sonication. After centrifugation, the clarified lysate was added to pre-equilibrated glutathione Sepharose 4B resin beads (GE Healthcare) and incubated at 4°C for 1 hour with end-over-end rotation. The beads were then washed 4 times with a buffer containing cold PBS and eluted with a buffer containing 50 mM Tris pH 8.0 and 10 mM reduced glutathione. After SDS-PAGE analysis, the purified protein was concentrated with a 30,000 Da molecular weight cut-off (MWCO) spin filter (Millipore), buffer exchanged into PBS and 50% glycerol, and stored at -20°C.

Expression and Purification of Sult1a1, Chst11, and Ust. BL21(DE3)RIL (Stratagene) harboring the pET28a-Sult1a1 construct and Origami B(DE3) (Novagen) harboring the

pET32b-hChst1₁₇₂₋₃₅₂ or pET32b-hUst₈₄₋₄₀₆ constructs were kindly provided by Dr. Jian Liu (UNC-Chapel Hill). Cells were grown in LB medium (1L) at 37°C. When cells reached an OD₆₀₀ of 0.8, cells were moved to 22°C for 30 minutes. Isopropyl β -D-1-thiogalactopyranoside (IPTG) was added to the cells at the final concentrations of 0.2 mM after the 30 minutes and cells were incubated for an additional 18 hours at 22°C. The pelleted cells were lysed in a cold buffer containing 25 mM Tris pH 7.5, 500 mM NaCl, 10 mM imidazole and 1x EDTA-free CompleteTM protease inhibitors (Roche) by sonication. After centrifugation, the clarified lysate was added to pre-equilibrated Ni-NTA resin beads (Qiagen) and incubated at 4°C for 1 hour with end-over-end rotation. The beads were then washed 3 times with a buffer containing 25 mM Tris pH 7.5, 500 mM NaCl, 10 mM imidazole, washed 1 time with 25 mM Tris pH 7.5, 500 mM NaCl, 50 mM imidazole and eluted with a buffer containing 25 mM Tris pH 7.5, 500 mM NaCl, 250 mM imidazole. After SDS-PAGE analysis, the purified protein was concentrated with a 30,000 Da molecular weight cut-off (MWCO) spin filter (Millipore), buffer exchanged into 25 mM Tris pH 7.5, 500 mM NaCl and 50% glycerol, and stored at -20°C.

96-Well Plate Enzyme-Coupled, Fluorescent Assay. For full assay, the substrates PAPS and CS-A (Sigma) were added to a solution containing 100 mM Tris pH 7.6, 1 mM β -mercaptoethanol, and 5 mM 4-methylumbelliferyl sulfate at a concentration of 30 μ M and 1 mM, respectively, to a total volume of 74 μ L. 5 μ L of Sult1c1 (2 mg/mL) was added to the mixture and incubated at 37°C for 1 hour. 20 μ L of Chst15 (1 mg/mL), pre-incubated with 1 μ L of DMSO or inhibitor stock for 30 minutes, was then added to the mixture to initiate the reaction. The 96-well plate was read using either a Victor3 Plate Reader (Perkin

Elmer) or Flexstation 3 (Molecular Devices) at excitation and emission wavelengths of 360 nm and 449 nm respectively.

For the counter screen, the substrates PAPS and CS-A (Sigma) were also added to a solution containing 100 mM Tris pH 7.6, 1 mM β -mercaptoethanol, and 5 mM 4-methylumbelliferyl sulfate at a concentration of 30 μ M and 1 mM, respectively, to a total volume of 75 μ L. The reaction was initiated by adding a mixture of 5 μ L of Sult1c1, 19 μ L of 100 mM Tris pH 7.6, and 1 μ L of DMSO or inhibitor stock that has been pre-incubated at room temperature for 30 minutes.

For enzyme specificity determination, Chst15 was replaced with either Chst11, Ust, Hs3st1, Sult1e1, Sult2b1a, or Sult2b1b. Likewise, CS-A was replaced with unsulfated chondroitin (Seikagaku), CS-C (Sigma), heparan sulfate (Neoparin), β -estradiol (Sigma), 5-pregnen-3 β -ol-20-one (Sigma), or cholesterol (Sigma) and used with their respective enzymes at 1 mM. IC₅₀'s were determined by varying the inhibitor concentration and fit to a Michaelis-Menten model by Kaleidagraph (version 4.1.2).

For Chst15 kinetic analysis, reactions were performed in triplicate with varying substrate PAPS (10 μ M to 200 μ M) or CS-A (0.5 mM to 10 mM) concentrations. Product formation was monitored for 10 minutes and kinetic parameters were determined from Lineweaver-Burk, K_m/V versus [inhibitor], and 1/V versus [inhibitor] plots.

Cell Culture. NIH/3T3 fibroblasts and Neu7 astrocytes were grown in DMEM medium supplemented with 10% fetal bovine serum (FBS), 100 units/mL penicillin, and 0.1 mg/mL streptomycin (Gibco). Cells were passaged and dissociated with 0.25% trypsin (Life

Technologies). Cells were incubated in a 5% CO₂ humidified chamber at 37°C. NIH/3T3 fibroblasts were obtained from American Type Culture Collection (ATCC) and Neu7 astrocytes were generously gifted by Dr. Herbert Geller (NIH).

CS/HS Expression by Immunofluorescence. NIH/3T3 fibroblasts or Neu7 astrocytes were seeded at 4,000 cells/cm² and grown in media premixed with inhibitors diluted from a 100x DMSO stock solution. After 48 hours, the media was aspirated and cells were fixed with 4% paraformaldehyde (PFA) for 15 minutes at room temperature. For chondroitinase ABC (ChABC) background determination, cells were treated with ChABC (Seikagaku) at a final concentration of 2 mU/μL in 33 mM Tris, 33 mM NaOAc pH 8.0 for 20 minutes at 37°C and 5% CO₂ prior to fixing. For heparinase background determination, cells were treated with heparinase I and III (Sigma) and II (Sigma) at final concentrations of 1 mU/μL and 0.2 mU/μL respectively in 50 mM NaPi pH 7.0 for 20 minutes at 37°C and 5% CO₂ prior to fixing. Cells were rinsed twice with PBS and blocked with a solution containing 1% BSA and 0.1% Triton X-100 in PBS for 30 minutes at room temperature. The primary antibody in a buffer containing 1% BSA and 0.1% Triton X-100 in PBS was then applied overnight at 4°C. For CS-E staining, a monoclonal mouse anti-CS-E antibody (1:250 dilution) was used. For HS staining, a monoclonal mouse heparan sulfate antibody (Amsbio, 10E4, 1:100 dilution) was used. After, cells were again washed twice with PBS and incubated with an anti-mouse AF488 antibody ((Life Technologies, 1:1000 dilution) in a buffer containing 1% BSA and 0.1% Triton X-100 in PBS for 1 hour at room

temperature. Cells were washed twice with PBS, imaged under a Zeiss 700 confocal microscope, and quantified by ImageJ.

CSPG Purification. Media from Neu7 astrocytes was collected 48 hours after treatment. The media was centrifuged at 4,000 x g for 10 minutes to remove cells and debris. Urea (2 M, final) was added to the supernatant and incubated at 4°C for 1 hour. For each 10 mL of media, 2 mL of pre-equilibrated DEAE Sepharose beads were added to the supernatant and incubated at 4°C for 1 hour. The mixture was then added to a column and allowed to drain. The beads were then washed with a buffer containing 50 mM Tris pH 7.5, 2 mM EDTA, 2 M urea, and 0.25 M NaCl. The CSPGs were eluted with a buffer containing 50 mM Tris pH 7.5, 2 mM EDTA, 2 M urea, and 2 M NaCl. The eluate was then dialyzed overnight at 4°C into PBS using dialysis tubing with a 10,000-12,000 MWCO (Spectrum Labs). The samples were then concentrated using 50 kDa MWCO spin filters (Millipore) and uronic acid concentrations of the samples was determined by the carbazole assay.

Disaccharide Analysis. CSPGs purified from Neu7 astrocytes were dialyzed into a buffer containing 100 mM Tris pH 7.5 using a spin filter with a 3,000 MWCO (Millipore). 150 µL of a buffer containing 100 mM Tris pH 7.5, 0.5% SDS, and 10 mM CaCl₂ was added to the CSPGs (50 µL). Pronase (Roche) was added to a final concentration of 2 mg/mL and the reaction was incubated at 37°C for 18 hours. The samples were boiled for 10 minutes and centrifuged at 12,000 x g for 10 minutes at room temperature. The supernatant was then added to a DEAE Sepharose (500 µL) column, washed 3 times with 1 mL of a buffer

containing 50 mM Tris pH 7.5, 2 mM EDTA, 2 M urea, and 0.25 M NaCl, and eluted with 1.5 mL of a buffer containing 2 mM EDTA, 2 M urea, and 2 M NaCl. The eluate was buffer exchanged into a buffer containing 33 mM Tris and 33 mM NaOAc pH 8.0 using a spin filter with a 3,000 MWCO. The CSPGs (50 μ L) were diluted to 200 μ L with the same buffer. 50 mU of chondroitinase ABC (Seikagaku) was added to each CSPG sample and allowed to incubate at 37°C for 18 hours. The reaction was put into a 3,000 MWCO spin filter, centrifuged at 12,000 x g for 10 minutes and flowthrough was collected. This was repeated twice more by adding 200 μ L of water to the retentate. The flowthroughs were pooled, analyzed by the carbazole assay, and lyophilized. For 2-aminoacridone (AMAC) conjugation, 5 μ L of a solution containing 0.1 M AMAC in 3:18 glacial acetic acid:DMSO was added to 1 μ g of CS disaccharide and incubated for 15 minutes at room temperature. 5 μ L of 1 M NaBH₃CN was then added and the reaction was incubated for 4 hours at 45°C. The reaction was quenched with 1:1 DMSO:water and analyzed by HPLC with a Poroshell 120 EC-C18 column (4.6 x 50 mm) with the following method at a flow rate of 1 mL/min and detection at 428 nm: A linear gradient of 98% 60 mM NH₄OAc and 2% MeCN to 70% 60 mM NH₄OAc and 30% MeCN for 50 minutes, followed by 15 minutes of 98% 60 mM NH₄OAc and 2% MeCN.

Neurite Outgrowth Assays. For outgrowth in inhibitor studies, 96-well poly-lysine Cellware plates (BD BioCoat™) were coated with laminin (10 μ g/mL, Life Technologies) in Neurobasal medium (Life Technologies) for 2 hours at 37°C and 5% CO₂. Dorsal root ganglia neurons (DRGs) were dissected from P5 Sprague Dawley rats, incubated in 0.25%

trypsin with EDTA for 15 minutes at 37°C, followed by collagenase (Worthington; 4 mg/mL) for 15 minutes at 37°C, triturated to dissociate to single cell suspensions, filtered using a 40 µm cell strainer (Fisher) to remove nondissociated cells, and seeded at approximately 60,000 cells/cm² in Neurobasal medium supplemented with B27 and GlutaMAXTM (Life Technologies). After 48 hours, the media was aspirated and cells were fixed with 4% paraformaldehyde (PFA) for 15 minutes at room temperature. A monoclonal rabbit anti-β3 tubulin antibody (1:1000 dilution, Cell Signaling (D71G9)) in a buffer containing 1% BSA and 0.1% Triton X-100 in PBS was then applied overnight at 4°C. After, cells were again washed twice with PBS and incubated with an anti-rabbit AF568 antibody ((Life Technologies, 1:1000 dilution) in a buffer containing 1% BSA and 0.1% Triton X-100 in PBS for 1 hour at room temperature. Cells were washed twice with PBS, imaged under a Zeiss 700 confocal microscope, and quantified using MetaMorph software.

For inhibition studies using CSPGs derived from Neu7 astrocytes, glass coverslips were coated with poly-lysine in borate buffer pH 8.5 (0.5 mg/mL) overnight at 37°C and 5% CO₂, followed by laminin (10 µg/mL) in Neurobasal medium (Life Technologies) for 2 hours at 37°C and 5% CO₂, followed by CSPGs purified from Neu7 astrocytes (1 µg/mL based on uronic acid content) in PBS for 2 hours at 37°C and 5% CO₂. DRGs were dissected, plated, imaged, and quantified as above. *P*-values were determined using one-way ANOVA (n = 20-80 neurons per experiment) and the results from three experiments were shown.

Boundary Assay. CSPGs purified from Neu7 astrocytes (5 µg/mL based on uronic acid content) were mixed with Texas Red (0.5 mg/mL; Life Technologies) in PBS, spotted on

poly-lysine and laminin (10 µg/mL) coated coverslips (3 µL), and incubated for 2 hours at 37°C and 5% CO₂. DRGs were dissected from P5 Sprague Dawley rats, incubated in 0.125% trypsin with EDTA for 15 minutes at 37°C, followed by collagenase (Worthington; 4 mg/mL) for 15 minutes at 37°C, triturated to dissociate to single cell suspensions, filtered using a 40 µm cell strainer (Fisher) to remove nondissociated cells and seeded at approximately 60,000 cells/cm². Cells were cultured for 48 hours in Neurobasal medium supplemented with B27 and GlutaMAXTM (Life Technologies). After immunostaining for neurite outgrowth, axons growing toward and within 10 µm of the boundary were evaluated using a Zeiss 700 confocal microscope. The percentage of axons that crossed the boundary over the total axons was quantified. *P*-values were determined using one-way ANOVA (*n* = 20-40 neurons per experiment) and the results from five experiments were shown.

REFERENCES

- (1) Kobayashi, M.; Habuchi, H.; Yoneda, M.; Habuchi, O.; Kimata, K. *J. Biol. Chem.* **1997**, 272, 13980.
- (2) Aikawa, J.; Esko, J. D. *J. Biol. Chem.* **1999**, 274, 2690.
- (3) Aikawa, J.; Grobe, K.; Tsujimoto, M.; Esko, J. D. *J. Biol. Chem.* **2001**, 276, 5876.
- (4) Habuchi, H.; Kobayashi, M.; Kimata, K. *J. Biol. Chem.* **1998**, 273, 9208.
- (5) Habuchi, H.; Tanaka, M.; Habuchi, O.; Yoshida, K.; Suzuki, H.; Ban, K.; Kimata, K. *J. Biol. Chem.* **2000**, 275, 2859.
- (6) Yamauchi, S.; Mita, S.; Matsubara, T.; Fukuta, M.; Habuchi, H.; Kimata, K.; Habuchi, O. *J. Biol. Chem.* **2000**, 275, 8975.
- (7) Okuda, T.; Mita, S.; Yamauchi, S.; Fukuta, N.; Nakano, H.; Sawada, T.; Habuchi, O. *J. Biol. Chem.* **2000**, 275, 40605.

- (8) Fukuta, M.; Inazawa, J.; Torii, T.; Tsuzuki, K.; Shimada, E.; Habuchi, O. *J. Biol. Chem.* **1997**, 272, 32321.
- (9) Bistrup, A.; Bhakta, S.; Lee, J. K.; Belov, Y. Y.; Gunn, M. D.; Zuo, F. R.; Huang, C. C.; Kannagi, R.; Rosen, S. D.; Hemmerich, S. *J. Cell Biol.* **1999**, 145, 899.
- (10) Hiraoka, N.; Petryniak, B.; Nakayama, J.; Tsuboi, S.; Suzuki, M.; Yeh, J. C.; Izawa, D.; Tanaka, T.; Miyasaka, M.; Lowe, J. B.; Fukuda, M. *Immunity* **1999**, 11, 79.
- (11) Krishnan, S.; Miller, R. M.; Tian, B.; Mullins, R. D.; Jacobson, M. P.; Taunton, J. *J. Am. Chem. Soc.* **2014**, 136, 12624.
- (12) Serafimova, I. M.; Pufall, M. A.; Krishnan, S.; Duda, K.; Cohen, M. S.; Maglathlin, R. L.; McFarland, J. M.; Miller, R. M.; Frodin, M.; Taunton, J. *Nat. Chem. Biol.* **2012**, 8, 471.
- (13) Brown, J. M.; Xia, J.; Zhuang, B. Q.; Cho, K. S.; Rogers, C. J.; Gama, C. I.; Rawat, M.; Tully, S. E.; Uetani, N.; Mason, D. E.; Tremblay, M. L.; Peters, E. C.; Habuchi, O.; Chen, D. F.; Hsieh-Wilson, L. C. *P. Natl. Acad. Sci. U.S.A.* **2012**, 109, 4768.
- (14) Gilbert, R. J.; McKeon, R. J.; Darr, A.; Calabro, A.; Hascall, V. C.; Bellamkonda, R. V. *Mol. Cell Neurosci.* **2005**, 29, 545.
- (15) Baeuerle, P. A.; Huttner, W. B. *Biochem. Biophys. Res. Co.* **1986**, 141, 870.
- (16) Steffen, C.; Wetzel, E. *Toxicology* **1993**, 84, 217.
- (17) Fok-Seang, J.; Smith-Thomas, L. C.; Meiners, S.; Muir, E.; Du, J. S.; Housden, E.; Johnson, A. R.; Faissner, A.; Geller, H. M.; Keynes, R. J.; Rogers, J. H.; Fawcett, J. W. *Brain Res.* **1995**, 689, 207.
- (18) Laabs, T. L.; Wang, H.; Katagiri, Y.; McCann, T.; Fawcett, J. W.; Geller, H. M. *J. Neurosci.* **2007**, 27, 14494.
- (19) Volpi, N.; Galeotti, F.; Yang, B.; Linhardt, R. J. *Nat. Protoc.* **2014**, 9, 541.
- (20) Fawcett, J. W.; Asher, R. A. *Brain Res. Bull.* **1999**, 49, 377.
- (21) Chin, P. C.; Liu, L.; Morrison, B. E.; Siddiq, A.; Ratan, R. R.; Bottiglieri, T.; D'Mello, S. R. *J. Neurochem.* **2004**, 90, 595.

Chapter 5

INTRODUCTION TO *O*-GLCNAC AND STUDYING ITS ROLE IN STRUCTURE AND FUNCTION

Chapter 5

INTRODUCTION TO *O*-GLCNAC AND STUDYING ITS ROLE IN STRUCTURE AND FUNCTION

The addition of *O*-linked β -*N*-acetylglucosamine (*O*-GlcNAc) to serine or threonine residues of proteins is a key regulator of many physiological functions, ranging from nutrient sensing^{1,2} and insulin signaling^{3,4} to learning and memory.^{5,6} Post translational modification of proteins with *O*-GlcNAc is a unique form of glycosylation that occurs intracellularly and dynamically, akin to protein phosphorylation. Over 1000 proteins bearing the modification have been identified to date, ranging from kinases, structural proteins, and nuclear pore proteins to transcriptional machinery including RNA polymerase II.⁷ Dysregulation of the modification has also been linked to disease states such as cancer,^{4,8,9} diabetes,^{4,10} and Alzheimer's.^{11,12} Despite these important implications, the precise roles of the modification are only beginning to be understood. Major challenges to the study of *O*-GlcNAc include its difficulty in detection, substoichiometric presence and ubiquitous nature, often modifying multiple sites on the same protein. As a result, detailed mechanistic studies have been lacking in the field, and only a few studies have established how site-specific glycosylation affects protein function. With that said, the development of chemical approaches to study the site-specific effects of *O*-GlcNAcylation offer a route to overcome these barriers of study. Using these methods to synthesize homogeneously glycosylated proteins will be critical to further understanding the role this modification plays in protein structure and function.

Protein Modification by O-GlcNAc

Unlike the 518 kinases and ~200 phosphatases that phosphorylate and dephosphorylate different protein substrates, the covalent modification of *O*-GlcNAc is regulated by two enzymes, *O*-GlcNAc transferase (OGT) and *O*-GlcNAcase (OGA) (Figure 5.1).¹³⁻¹⁶ OGT and OGA catalyze the reversible addition and removal of *O*-GlcNAc respectively. How these two enzymes append and remove *O*-GlcNAc site-specifically on particular protein substrates is largely unknown. Activity of these two enzymes is thought to be regulated by accessory proteins that potentiate their specificity. For example, the three OGT isoforms differ in the number of tetratricopeptide repeats (TPRs) within its N-terminal domain.¹⁷ These TPRs serve as protein-protein interaction modules that have been shown to interact with proteins such as the GABA_A receptor-associated protein (GRIF-1) and its homolog, *O*-GlcNAc transferase interacting protein (OIP106).¹⁸

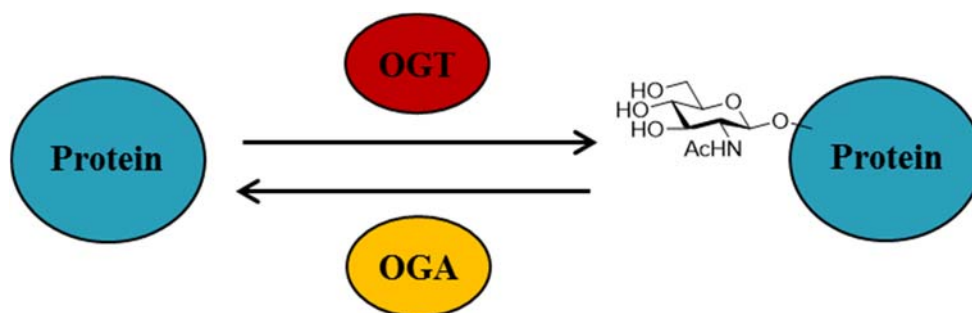


Figure 5.1. Dynamic, covalent attachment and removal of *O*-GlcNAc is catalyzed by two enzymes, *O*-GlcNAc transferase (OGT) and *O*-GlcNAcase (OGA), respectively.

From a functional perspective, *O*-GlcNAc has been shown to play roles in both disrupting^{5,19,20} and enhancing²¹⁻²³ protein-protein interactions as well as competing with phosphorylation directly.^{19,24,25} In particular, this has led to interesting studies and questions on the interplay between phosphorylation and glycosylation.^{19,26,27} In these studies, a global increase in *O*-GlcNAcylation was shown to change the stoichiometry of almost all presently cycling phosphorylation sites among 711 phosphopeptides. However, general analyses the site specific effects and interplay of *O*-GlcNAc has been severely hampered by the difficulty in detection of the modification, chemical and enzymatic lability, substoichiometric presence, and complex cellular regulation. The field would greatly benefit from the use of techniques to overcome these limitations and dissect the functional implications of the modifications.

Cyclic-AMP Response Element Binding Protein (CREB)

Cyclic-AMP response element binding protein (CREB) is a transcription factor that underlies diverse processes, ranging from long-term potentiation and drug addiction to circadian rhythm. It has been extensively studied in the context of its inducible activity through many cellular stimuli, including cyclic-AMP (cAMP) and Ca²⁺ signaling, leading to the attachment of posttranslational modifications. Phosphorylation of Ser133, the most well studied regulatory site on the protein, leads to the recruitment of the coactivator CREB-binding protein (CBP) and activation of CREB-mediated transcription. However, there also exist other mechanisms for CREB-dependent transcription as phosphorylation is now always sufficient to stimulate activity.

Work in our lab has shown that *O*-GlcNAc can serve as another mechanism for gene regulation through the modification of CREB.^{5,28} CREB was found to contain six potential sites of glycosylation and each was variably modified. Our data suggested that the predominant glycosylation site, Ser40, served as a constant repressor of CREB transcriptional activity, lowering basal levels of transcription. In addition to modulating constitutive transcription, observed increases in glycosylation after neuronal stimulation were proposed to be part of a larger complex signaling mechanism to balance gene transcription. *In vivo* expression of an S40A CREB mutant led to mice with increased capacity to form long term memory. Interestingly, on a molecular level, glycosylation at Ser40 disrupts the binding of phosphorylation-independent CREB-regulated transcription coactivator (CRTC), known to bind to the bZIP domain.²⁹ How a small modification near the N-terminus can effect a protein-protein interaction by the C-terminus is unknown, but implicates a long-range interaction between the ends of this protein.

Structurally, CREB is made up of four domains: two glutamine rich domains (Q1 and Q2) separated by the kinase inducible domain (KID) and a C-terminal bZIP domain (Figure 5.2).³⁰ Each domain serves to facilitate interactions with coactivators and components of the transcriptional machinery. Despite the importance of CREB and its discovery thirty years ago, the complete structure of the protein has yet to be solved.³¹ It has been hypothesized that the protein is disordered making structural studies challenging. Nevertheless, the structure of two of the four domains of CREB have been solved. The bZIP domain of CREB bound to a CRE oligonucleotide has been previously crystallized. The structure showed the important interactions that allow for CRE specific binding and

dimerization between CREB family members.³² Key residues conserved between bZIP containing family members such as Arg301 and Lys304 make central interactions with the core base pairs that define the CRE sequence. Furthermore, The solution structure of the interaction between the CREB KID domain and CBP KIX has been previously solved.³³ This structure elucidated the important interactions between CREB and CBP required for phosphorylation-dependent transcription. Phosphorylation at Ser133 of the KID domain allows for the formation of a critical hydrogen bond anchor between the phosphate oxygen and Tyr658 hydroxyl in the KIX domain. Mutation of Ser133 to alanine abolished their interaction *in vitro*. Importantly, these studies serve as precedence for the induction of structural changes following the addition of post-translational modifications on CREB. These studies warrant the investigation of how *O*-GlcNAc may be effecting CREB structure and transcriptional activity.

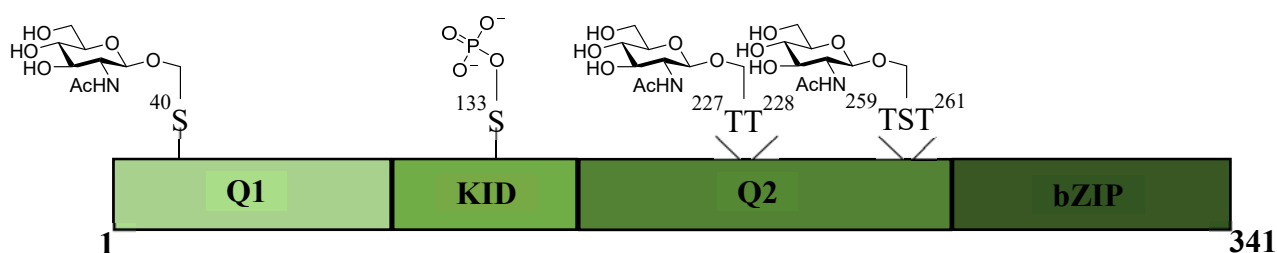


Figure 5.2. Four domains of CREB and locations of key phosphorylation and *O*-GlcNAcylation sites. CREB is composed of four domains consisting of glutamine rich domains (Q1 and Q2), kinase inducible domain (KID), and a basic leucine zipper (bZIP) domain. Phosphorylation at Ser133 has been shown to be a key regulatory element in the activation of CREB transcription. Six *O*-GlcNAc modified sites have been identified throughout the Q1 and Q2 domains with Ser40 being the predominant site of glycosylation.

Structural Implications of O-GlcNAc

The effects of O-GlcNAc at the protein level are unknown (e.g. steric hindrance, creation of a protein-binding surface, conformational change, etc.). Studies on the protein level have been hindered by the difficulty in acquiring homogeneous and site-specifically glycosylated proteins. However, there exist studies on the peptide level that have provided insights into the potential role O-GlcNAc on protein structure.

Wong and coworkers found that when the threonine residue of the polymeric domain of RNA polymerase II, a known site of glycosylation, is *O*-GlcNAcylated, a turn-like structure is induced when compared to the random coil unglycosylated peptide.³⁴ In another study by Wong and coworkers, the conformational changes of a peptide of the mucin domain of MAdCAM-1 were examined by NMR and computational modeling when different carbohydrates, including *O*-GlcNAc, were attached. The NMR data showed that the *O*-GlcNAc peptide was clearly to interact with neighboring sidechains. Upon further elaboration with galactose and sialic acid, these noticeable GlcNAc interactions disappeared, implying a conformational change.³⁵ Li and coworkers examined similar changes by NMR when using the N-terminus peptide of murine estrogen receptor β . Attachment of an *O*-GlcNAc moiety created a higher degree of the type II β -turn-like structure in the region. Interestingly, adding a phosphorylation at the same Ser16 residue induced a more extended structure.³⁶

While these studies implicate the potential for *O*-GlcNAc to induce structural change, they model the change on short peptides and do not exclude other mechanisms of by which *O*-GlcNAc can effect function. Expansion of such analyses to proteins will be

essential to understand the effects of O-GlcNAcylation on protein structure and dynamics, including its potential to modulate long-range interactions.

Methods for the Semi-Synthesis of O-GlcNAcylated Proteins

Use of chemical approaches presents one of the few avenues in which the structural effects of site-specific glycosylation can be studied. There exists a limited amount of techniques to generate homogeneously glycosylated protein. The two principal methods for achieving this are expressed protein ligation and bioconjugation of an introduced dehydroalanine (Dha). These methods have given researchers great insights into the functional roles these proteins play, but have yet to bear any structural information.

Expressed protein ligation has proven to be a powerful tool in the generation of semi-synthetic proteins.³⁷⁻³⁹ This approach utilizes the unique reactivity of a protein/peptide fragment containing an N-terminal cysteine with another protein/peptide fragment containing a C-terminal thioester. The method can be applied universally to almost any protein containing cysteines or tolerant of the introduction of cysteines. There also exist methods of radical desulfurization to convert cysteine to alanine, expanding the scope of this method.⁴⁰ The limitations of this method derives from limiting the number of ligations as protein yield and purity decrease with increasing ligations. In the context of O-GlcNAc, there are two examples in which this form of semisynthesis has been used to elucidate the biological implications of O-GlcNAc. Cole and coworkers used a one-step ligation to study the effects of O-GlcNAcylation and phosphorylation on casein kinase 2 (CK2). They determined that phosphorylation at Thr344 and glycosylation at Ser347

played critical roles in determining the substrate specificity of CK2.⁴¹ In another example, Pratt and coworkers utilized a two-step ligation to generate α -synuclein *O*-GlcNAcylated at Thr72. The functional consequence of this modification was determined to be an increase in solubility and decrease in the number of aggregates that are usually related to neurodegeneration.⁴²

The method developed by Davis and coworkers is based on the facile alkylation and elimination of a cysteine sidechain to Dha.⁴³⁻⁴⁶ The reaction has been accomplished by the development of two reagents: *O*-mesitylenesulfonylhydroxylamine (MSH) and 2,5-dibromohexanediamide (DBHDA). This transformation provides an electrophilic handle for the nucleophilic addition of different kinds of thiol-containing post-translation modification analogs (Figure 5.3). This method has been successfully demonstrated for the attachment of phosphorylation, glycosylation, acetylation, and methylation mimics. Despite the likeliness of forming C α epimers at the amino acid of interest, the proteins that have been semi-synthesized through this approach have been functionally active.⁴³⁻⁴⁶ Using this method, Davis and coworkers were able to show that these mimics were recognized by antibodies developed for the natural modifications, as well as successful deacetylation of the acetyllysine mimic by histone deacetylase (HDAC) 1 and HDAC2.⁴⁴ They also used this method to support the hypothesis that *O*-GlcNAc glycosylation of histone H2A at Thr101 disrupts nucleosome formation and increases transcriptional activity by lowering the barrier of RNA polymerase passage.⁴⁶ The largest limitation for this method is that it requires other cysteine residues to be mutated to their serine counterparts. If the protein

cannot tolerate this change (i.e. disulfide bridges, catalytic cysteines, etc.) then the method is most likely not amenable.

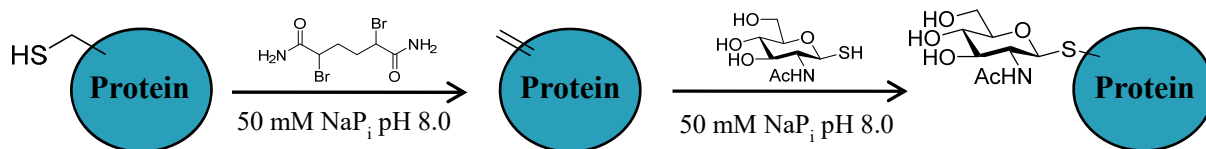


Figure 5.3. One-pot chemical glycosylation reaction through the conversion of dehydroalanine (Dha). Under slightly basic, buffering conditions, cysteine sidechains can be alkylated and eliminated with the addition of 2,5-dibromohexanediamide (DBHDA). The following excess addition of a thiol-containing compound leads to the generation of homogeneously glycosylated protein.

REFERENCES

- (1) Zachara, N. E.; Hart, G. W. *Bba-Gen. Subjects* **2004**, 1673, 13.
- (2) Hardiville, S.; Hart, G. W. *Cell Metab.* **2014**, 20, 208.
- (3) Yang, X. Y.; Ongusaha, P. P.; Miles, P. D.; Havstad, J. C.; Zhang, F. X.; So, W. V.; Kudlow, J. E.; Michell, R. H.; Olefsky, J. M.; Field, S. J.; Evans, R. M. *Nature* **2008**, 451, 964.
- (4) Slawson, C.; Copeland, R. J.; Hart, G. W. *Trends Biochem. Sci.* **2010**, 35, 547.
- (5) Rexach, J. E.; Clark, P. M.; Mason, D. E.; Neve, R. L.; Peters, E. C.; Hsieh-Wilson, L. C. *Nat. Chem. Biol.* **2012**, 8, 253.
- (6) Tallent, M. K.; Varghis, N.; Skorobogatko, Y.; Hernandez-Cuebas, L.; Whelan, K.; Vocadlo, D. J.; Vosseller, K. *J. Biol. Chem.* **2009**, 284, 174.
- (7) Sakabe, K.; Wang, Z. H.; Hart, G. *Glycobiology* **2009**, 19, 1368.
- (8) Ma, Z. Y.; Vosseller, K. *Amino Acids* **2013**, 45, 719.
- (9) Slawson, C.; Hart, G. W. *Nat. Rev. Cancer* **2011**, 11, 678.

- (10) Banerjee, P. S.; Ma, J. F.; Hart, G. W. *P. Natl. Acad. Sci. U.S.A.* **2015**, *112*, 6050.
- (11) Yuzwa, S. A.; Shan, X. Y.; Macauley, M. S.; Clark, T.; Skorobogatko, Y.; Vosseller, K.; Vocadlo, D. J. *Nat. Chem Biol.* **2012**, *8*, 393.
- (12) Liu, F.; Iqbal, K.; Grundke-Iqbal, I.; Hart, G. W.; Gong, C. X. *P. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 10804.
- (13) Manning, G.; Whyte, D. B.; Martinez, R.; Hunter, T.; Sudarsanam, S. *Science* **2002**, *298*, 1912.
- (14) Sacco, F.; Perfetto, L.; Castagnoli, L.; Cesareni, G. *FEBS Lett.* **2012**, *586*, 2732.
- (15) Hart, G. W.; Housley, M. P.; Slawson, C. *Nature* **2007**, *446*, 1017.
- (16) Lazarus, M. B.; Nam, Y. S.; Jiang, J. Y.; Sliz, P.; Walker, S. *Nature* **2011**, *469*, 564.
- (17) Lazarus, B. D.; Love, D. C.; Hanover, J. A. *Glycobiology* **2006**, *16*, 415.
- (18) Iyer, S. P. N.; Akimoto, Y.; Hart, G. W. *J. Biol. Chem.* **2003**, *278*, 5399.
- (19) Wang, Z.; Gucek, M.; Hart, G. W. *P Natl Acad Sci USA* **2008**, *105*, 13793.
- (20) Yang, W. H.; Kim, J. E.; Nam, H. W.; Ju, J. W.; Kim, H. S.; Kim, Y. S.; Cho, J. W. *Nat. Cell Biol.* **2006**, *8*, 1074.
- (21) Fujiki, R.; Hashiba, W.; Sekine, H.; Yokoyama, A.; Chikanishi, T.; Ito, S.; Imai, Y.; Kim, J.; He, H. H.; Igarashi, K.; Kanno, J.; Ohtake, F.; Kitagawa, H.; Roeder, R. G.; Brown, M.; Kato, S. *Nature* **2011**, *480*, 557.
- (22) Gewinner, C.; Hart, G.; Zachara, N.; Cole, R.; Beisenherz-Huss, C.; Groner, B. *J. Biol. Chem.* **2004**, *279*, 3563.
- (23) Fujiki, R.; Chikanishi, T.; Hashiba, W.; Ito, H.; Takada, I.; Roeder, R. G.; Kitagawa, H.; Kato, S. *Nature* **2009**, *459*, 455.
- (24) Kamemura, K.; Hayes, B. K.; Comer, F. I.; Hart, G. W. *J. Biol. Chem.* **2002**, *277*, 19229.
- (25) Cheng, X. G.; Cole, R. N.; Zaia, J.; Hart, G. W. *Biochemistry* **2000**, *39*, 11609.
- (26) Comer, F. I.; Hart, G. W. *J. Biol. Chem.* **2000**, *275*, 29179.
- (27) Hart, G. W.; Slawson, C.; Ramirez-Correa, G.; Lagerlof, O. *Annu. Rev. Biochem.* **2011**, *80*, 825.

- (28) Lamarre-Vincent, N.; Hsieh-Wilson, L. C. *J. Am. Chem. Soc.* **2003**, *125*, 6612.
- (29) Luo, Q.; Viste, K.; Urdy-Zaa, J. C.; Kumar, G. S.; Tsai, W. W.; Talai, A.; Mayo, K. E.; Montminy, M.; Radhakrishnan, I. *P. Natl. Acad. Sci. U.S.A.* **2012**, *109*, 20865.
- (30) Lonze, B. E.; Ginty, D. D. *Neuron* **2002**, *35*, 605.
- (31) Montminy, M. R.; Sevarino, K. A.; Wagner, J. A.; Mandel, G.; Goodman, R. H. *P. Natl. Acad. Sci. U.S.A.* **1986**, *83*, 6682.
- (32) Schumacher, M. A.; Goodman, R. H.; Brennan, R. G. *J. Biol. Chem.* **2000**, *275*, 35242.
- (33) Radhakrishnan, I.; PerezAlvarado, G. C.; Parker, D.; Dyson, H. J.; Montminy, M. R.; Wright, P. E. *Cell* **1997**, *91*, 741.
- (34) Simanek, E. E.; Huang, D. H.; Pasternack, L.; Machajewski, T. D.; Seitz, O.; Millar, D. S.; Dyson, H. J.; Wong, C. H. *J. Am. Chem. Soc.* **1998**, *120*, 11567.
- (35) Wu, W. G.; Pasternack, L.; Huang, D. H.; Koeller, K. M.; Lin, C. C.; Seitz, O.; Wong, C. H. *J. Am. Chem. Soc.* **1999**, *121*, 2409.
- (36) Chen, Y. X.; Du, J. T.; Zhou, L. X.; Liu, X. H.; Zhao, Y. F.; Nakanishi, H.; Li, Y. M. *Chem. Biol.* **2006**, *13*, 937.
- (37) Dawson, P. E.; Muir, T. W.; Clarklewis, I.; Kent, S. B. H. *Science* **1994**, *266*, 776.
- (38) Muir, T. W.; Sondhi, D.; Cole, P. A. *P. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 6705.
- (39) Muir, T. W. *Annu. Rev. Biochem.* **2003**, *72*, 249.
- (40) Wan, Q.; Danishefsky, S. J. *Angew. Chem. Int. Edit.* **2007**, *46*, 9248.
- (41) Tarrant, M. K.; Rho, H. S.; Xie, Z.; Jiang, Y. L.; Gross, C.; Culhane, J. C.; Yan, G.; Qian, J.; Ichikawa, Y.; Matsuoka, T.; Zachara, N.; Etzkorn, F. A.; Hart, G. W.; Jeong, J. S.; Blackshaw, S.; Zhu, H.; Cole, P. A. *Nat. Chem. Biol.* **2012**, *8*, 262.
- (42) Marotta, N. P.; Lin, Y. H.; Lewis, Y. E.; Ambroso, M. R.; Zaro, B. W.; Roth, M. T.; Arnold, D. B.; Langen, R.; Pratt, M. R. *Nat. Chem.* **2015**, *7*, 913.
- (43) Bernardes, G. J. L.; Chalker, J. M.; Errey, J. C.; Davis, B. G. *J. Am. Chem. Soc.* **2008**, *130*, 5052.
- (44) Chalker, J. M.; Lercher, L.; Rose, N. R.; Schofield, C. J.; Davis, B. G. *Angew. Chem. Int. Edit.* **2012**, *51*, 1835.

(45) Gamblin, D. P.; Garnier, P.; van Kasteren, S.; Oldham, N. J.; Fairbanks, A. J.; Davis, B. G. *Angew. Chem. Int. Edit.* **2004**, *43*, 828.

(46) Lercher, L.; Raj, R.; Patel, N. A.; Price, J.; Mohammed, S.; Robinson, C. V.; Schofield, C. J.; Davis, B. G. *Nat. Commun.* **2015**, *6*.

Chapter 6

EXAMINING THE ROLE OF *O*-GLCNAC ON CREB THROUGH SEMISYNTHESIS

Chapter 6

EXAMINING THE ROLE OF *O*-GLCNAC ON CREB THOROUGH SEMISYNTHESIS

Introduction

The β -*N*-acetyl-D-glucosamine (*O*-GlcNAc) post-translational modification plays a major role in many diseases such as cancer,¹⁻⁴ diabetes,⁴⁻⁶ and neurodegenerative disorders,⁷⁻⁹ but much is still unknown about its molecular-level influence on protein structure and function. Although post-translational modifications have been known to impart structural changes leading to functional outcomes, notably, no structures of *O*-GlcNAcylated proteins exist. The challenge of obtaining homogeneous glycoproteins bearing the GlcNAc sugar at defined sites has hindered the biophysical studies of how this modification can affect structure on the protein level. Here we have utilized a semisynthetic approach to generate a homogeneously *O*-GlcNAcylated form of cyclic-AMP response element binding protein (CREB) for structural and functional studies.

Results and Discussion

To site-specifically study the role of *O*-GlcNAc in protein structure and function, we would need the homogeneously glycosylated protein of interest. For the protein of interest, we chose to investigate the role of *O*-GlcNAc on CREB because of our established interest in the modification's ability to repress long-term potentiation and memory formation through this transcription factor.^{10,11} To generate a homogenous population of glycosylated CREB,

we opted to use the established chemical glycosylation approach developed by Davis and coworkers that utilizes the conversion of a cysteine sidechain to an electrophilic dehydroalanine (Dha).¹²⁻¹⁴ This olefin sidechain can then serve as a point of nucleophilic attachment for a thiol-containing GlcNAc analog. An intrinsic limitation to this method is that it cannot be used on proteins containing native cysteine residues that are critical for protein structure and function. While CREB contains four native cysteine residues, three (C300, C310, C337) exist in the bZIP domain and have been previously mutated out with no effect on its ability to be crystallized bound to the CRE consensus sequence.¹⁵ We reasoned that since the last cysteine residue (C90) was not involved in disulfide bond formation, mutation to serine would be nonperturbing.

The most critical component for these studies was to express and purify this mutant form of CREB to homogeneity. Using site directed mutagenesis, we designed our construct to contain S40C, C90S, C300S, C310S, and C337S mutations. This allowed for bioconjugation to occur only at the amino acid site of interest (S40). The expression and handling of CREB has been plagued by poor stability, co-purification with DNA, and dimerization with truncation products; all of which would present issues for potential structural studies. To generate the homogenous protein, we adapted a previously reported purification protocol that utilized denaturing and

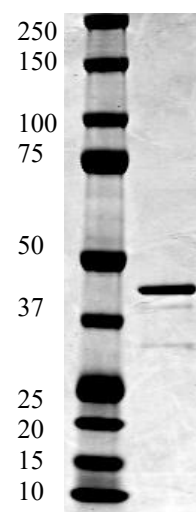


Figure 6.1. Mutant CREB was purified to >95% homogeneity through denaturing protocol as analyzed by Coomassie staining (lane 2).

refolding conditions to easily remove bound DNA and truncation products.¹⁶ In short, this method included heat extraction from lysate, denaturing nickel-affinity chromatography, hexahistidine-tag cleavage, protein refolding, and size exclusion chromatography (SEC) steps. Following these steps, we were able to obtain a protein of greater than 95% homogeneity (Figure 6.1). Interestingly, we observed that our mutant was stable for months at 4°C as opposed to the wild type protein, most likely a result of oxidation and aggregation of the free cysteine residues.¹⁷

To functionally verify that these procedures produced a refolded and biologically relevant protein, we tested its ability to bind the CRE consensus sequence and elicit a transcriptional response. DNA binding was assessed by electrophoretic mobility shift assay (EMSA) with an infrared-labeled CRE consensus sequence oligonucleotide. The binding of this mutant was confirmed and showed sequence selectivity as binding competed with an oligonucleotide bearing an identical sequence without the reporter (Figure 6.2A). To gauge if this purification process generated a transcriptionally active protein, an *in vitro* transcription assay, employing HeLa cell lysates, was used to detect activity. Natively expressed CREB, without denaturing purification, expectedly did not produce transcriptionally active CREB, most likely a result of being tightly bound to DNA. When purified under denaturing conditions, the mutant was unexpectedly 8-fold less active as compared to wild type CREB (Figure 6.2B). This decrease in activity, but not DNA binding, may be a result of diminished binding to the CRTC transcriptional cofactor as demonstrated with a similar triple mutant.¹⁸ Indeed, it was shown that the existence of C310 positively affects and C300 negatively regulates CREB-CRTC binding. However, we posited that as

long as there was sufficient signal to detect activity, we can still infer the effects of *O*-GlcNAcylation on CREB using this expressed form of the protein.

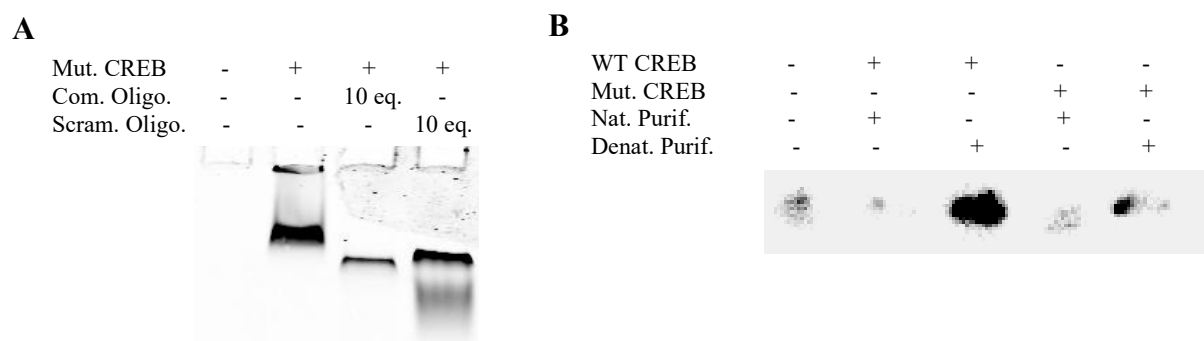


Figure 6.2. Expressed WT and mutant CREB were evaluated for their ability to bind DNA and lead to transcriptional activity. (A) The mutant CREB construct specifically bound to the CRE consensus sequence as analyzed by EMSA. Mut: mutant; Com: complementary; Scram: scramble; Oligo: oligonucleotide. (B) CREB purified natively showed no transcriptional activity. When purified under denaturing conditions, the mutant form was 8-fold less active in transcribing linearized DNA as compared to the WT protein. WT: wild type; Mut: mutant; Nat: native; Denat: denaturing; Purif: purification.

With the protein in hand, we sought to conjugate it to a GlcNAc analog. To achieve this, we synthesized the previously reported *O*-mesitylenesulfonylhydroxylamine (MSH) and 2,5-dibromohexanediamide (DBHDA) reagents in two steps each and have been previously reported to convert cysteine to Dha.^{12,19} The *S*-GlcNAc analog, 1-thio-2-acetamido-2-deoxy- β -D-glucopyranose was also synthesized in three steps to for the final conjugation step.¹³

We next optimized the conditions for the two-step conversion of CREB to *S*-GlcNAcylated CREB (Figure 6.3), by varying parameters such as the equivalents of the alkylating agent, protein concentration, reaction temperature, and time. Unexpectedly, we were unable to reproduce the ability of MSH to convert cysteine to Dha and thus moved on

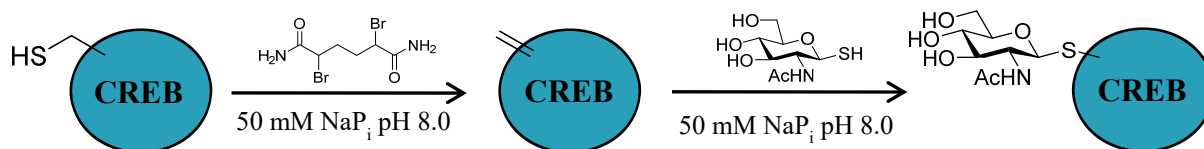


Figure 6.3. One-pot chemical glycosylation reaction of CREB. Under slightly basic, buffering conditions, cysteine sidechains can be alkylated and eliminated with the addition of 2,5-dibromohexanediamide (DBHDA). The following excess addition of a thiol-containing GlcNAc analog led to the generation of homogeneously glycosylated CREB.

Table 6.1. Cys to *S*-GlcNAc Conversion with DBHDA.

| Reaction Conditions | | | | | | | | Product Mass Detected ^a | Purity ^b (%) |
|-----------------------|--------------|-------------------|------------|-------------------|------------|---------------------|------------|------------------------------------|-------------------------|
| Protein Conc. (mg/mL) | DBHDA Equiv. | Dha Incubation #1 | | Dha Incubation #2 | | S-GlcNAc Incubation | | | |
| | | Time (min) | Temp. (°C) | Time (min) | Temp. (°C) | Time (min) | Temp. (°C) | | |
| 1 | 770 | 30 | RT | - | - | - | - | No | - |
| 1 | 770 | 30 | RT | 60 | 37 | - | - | No | - |
| 1 | 770 | 30 | RT | 90 | 37 | - | - | Yes | 27 |
| 1 | 770 | 30 | RT | 120 | 37 | - | - | Yes | 16 |
| 1 | 770 | 30 | RT | 18 h | RT | - | - | No | - |
| 1 | 770 | 30 | RT | 18 h | 4 | - | - | No | - |
| 1 | 10 | 30 | RT | 90 | 37 | - | - | No | - |
| 1 | 50 | 30 | RT | 90 | 37 | - | - | No | - |
| 1 | 100 | 30 | RT | 90 | 37 | - | - | Yes | 72 |
| 1 | 250 | 30 | RT | 90 | 37 | - | - | Yes | 73 |
| 1 | 100 | 15 | 37 | - | - | - | - | No | - |
| 1 | 100 | 30 | 37 | - | - | - | - | No | - |
| 1 | 100 | 60 | 37 | - | - | - | - | No | - |
| 1 | 100 | 90 | 37 | - | - | - | - | Yes | 33 |
| 2 | 100 | 30 | RT | 90 | 37 | - | - | Yes | 76 |
| 4 | 100 | 30 | RT | 90 | 37 | - | - | Yes | 77 |
| 1 | 770 | 30 | RT | 90 | 37 | 90 | RT | Yes | 26 |
| 1 | 100 | 30 | RT | 90 | 37 | 180 | RT | Yes | 44 |
| 1 | 100 | 30 | RT | 90 | 37 | 18 h | 4 | No | - |
| 2 | 100 | 30 | RT | 90 | 37 | 120 | 37 | Yes | 82 |
| 2 | 100 | 30 | RT | 90 | 37 | 18 h | RT | Yes | 75 |
| 2 | 100 | 30 | RT | 90 | 37 | 18 h | 37 | No | - |

^aCalculated Dha-containing CREB: 36607, GlcNAc-containing CREB: 36844. n.d.: not detected.

^bFinal purity was normalized to starting CREB purity after evaluation of LC-MS abundances

with the use of DBHDA. After initial reactions of CREB with DBHDA, we realized optimization of this first step was critical because of the observed likeliness to form many side products. To minimize these side products, iterative fine-tuning of these reaction conditions was performed and S40C CREB reacted with DBHDA (100 eq.) in pH 8.0 phosphate buffer for 30 min at room temperature then 37 °C for 1.5 h led to complete conversion to the corresponding Dha40 CREB, as the starting protein was no longer detected when analyzed by LC-MS (Table 6.1). Subsequent addition and incubation with *S*-GlcNAc (5000 eq.) for 2 h at 37 °C produced the desired *S*-GlcNAcylated product at a final purity of 82%, as evaluated by SDS-PAGE and LC-MS. (Figure 6.4). Importantly again, the Dha-containing protein mass was undetectable, indicating complete conversion after each step.

Further verification of the reaction was performed through collision-induced dissociation mass spectrometry (CID-MS) and purification by wheat germ agglutinin (WGA) affinity chromatography to enrich for GlcNAcylated CREB. In-gel trypsin digestion, followed by tandem MS, allowed us to confirm the presence of Dha and *S*-GlcNAc at the S40C position (Figure 6.5). The semi-synthetic protein was also specifically bound and eluted from solid support bound WGA, a lectin with a specific affinity for the GlcNAc moiety, indicating the presence of the chemical modification (Figure 6.6). It is important to note that the posttranslational modification mimics introduced by this method are likely formed as α epimeric mixtures. Nevertheless, despite this isomerism, these mimics have been shown to be functional in both immunoblot and enzymatic assays.^{14,19,20} Thus, it was still expected to generate functionally active proteins.

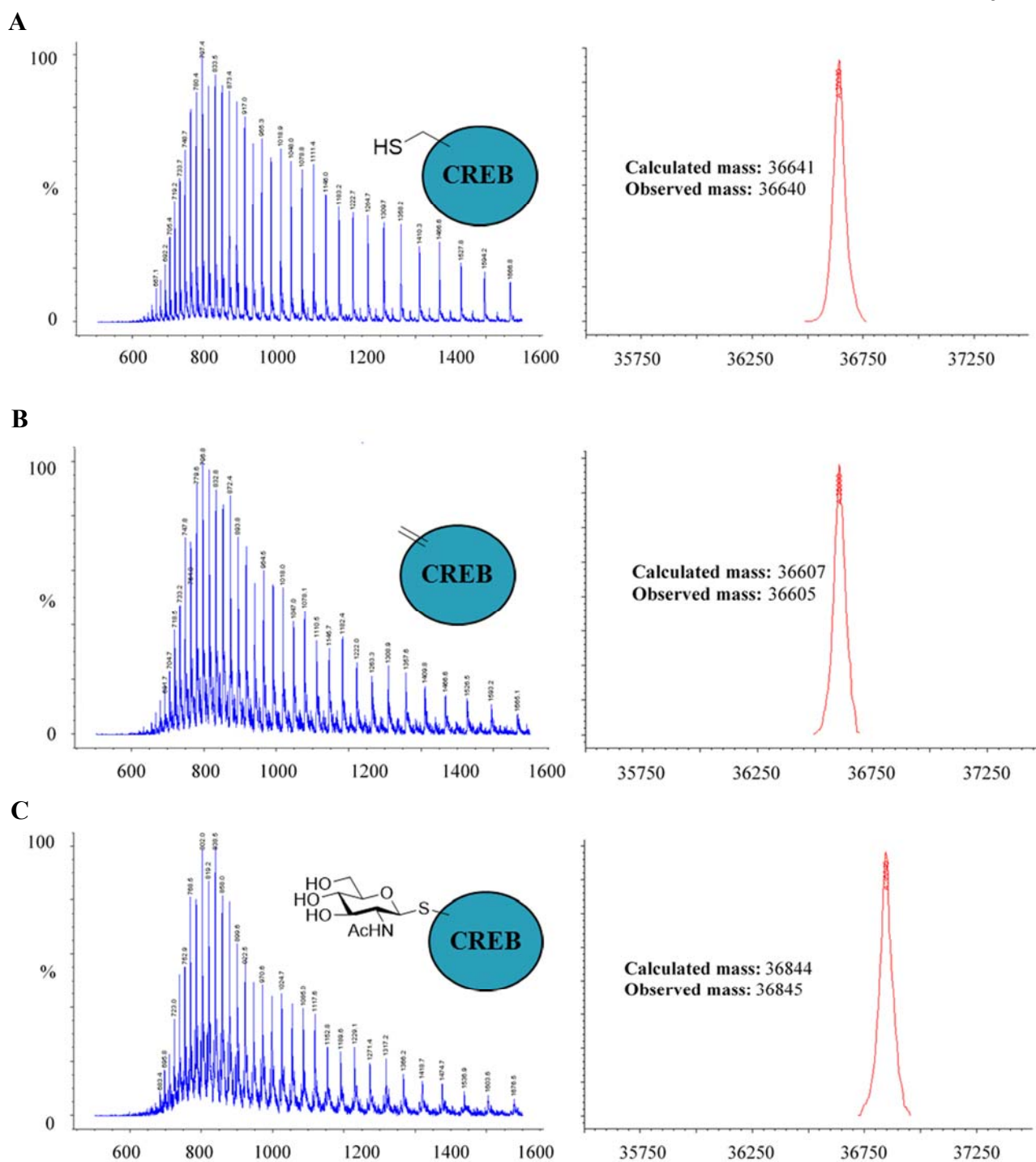


Figure 6.4. Intact protein LC-MS showed the rapid conversion of (A) mutant CREB to (B) Dha-containing CREB and to (C) S-GlcNAc-containing CREB. Mass shifts of -35 m/z and $+205$ m/z corresponded to the elimination of cysteine (Dha) and addition of S-GlcNAc, respectively. Full ion chromatogram (left) led to the expected deconvoluted (right) mass within error of the method. Deconvoluted mass spectra were analyzed with a 15% abundance cutoff with a minimum of 12 set peaks.

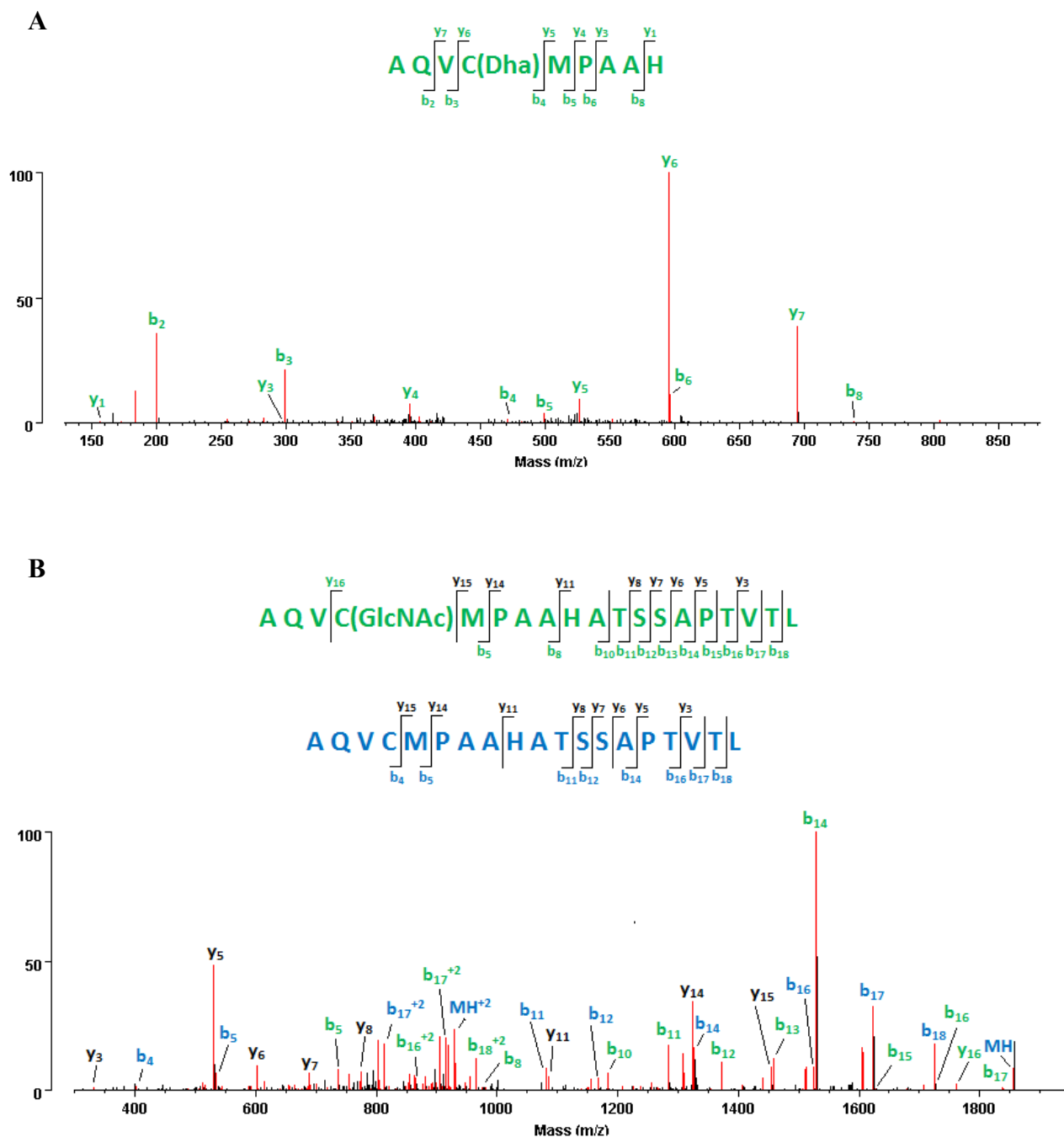


Figure 6.5. CID-MS analysis of in-gel chymotryptic digests after bioconjugation reactions identified CREB peptides bearing Dha and *S*-GlcNAc at S40C position. (A) Peptide containing Dha on S40C position after alkylation and elimination reaction. (B) Peptide fragments containing unique modified (green) and unmodified (blue) masses identified *S*-GlcNAc at S40C. Peptide fragments that are part of both modified and unmodified peptides are in black.

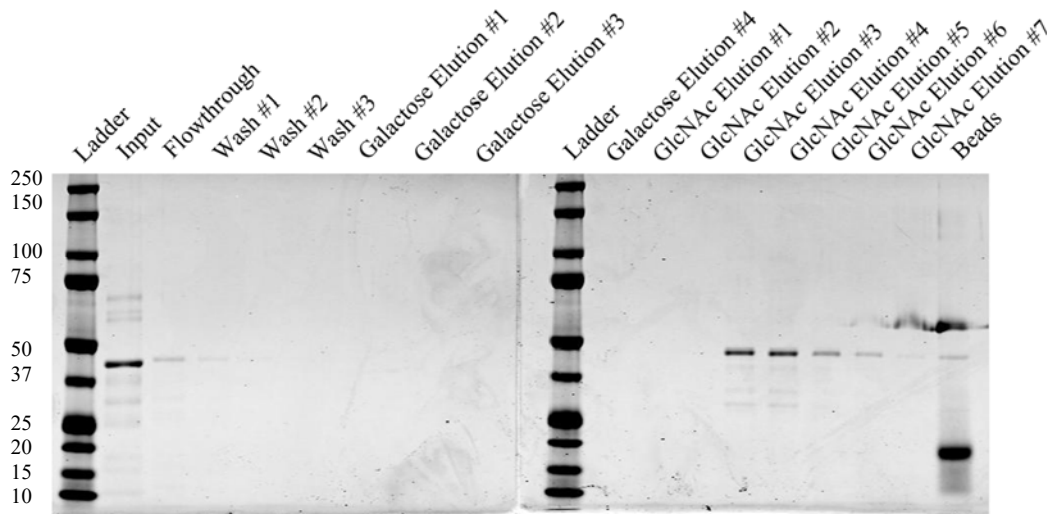


Figure 6.6. Wheat germ agglutinin (WGA) purification of semisynthetic CREB. WGA-bound agarose is able to recognize and specifically bind *S*-GlcNAcylated CREB as indicated by Coomassie staining. Minimal presence in the flowthrough and specific elution of CREB by GlcNAc support the success of the semisynthetic route.

To determine the integrity of the protein post-reaction, we assessed the ability of the protein to bind DNA. In cellular studies, we previously showed that CREB bearing a S40A mutation bound to the CRE consensus sequence with equal affinity as compared to the wild type.¹¹ Using EMSA, we surprisingly first observed no DNA binding by the reacted Dha and *S*-GlcNAc-containing proteins. However, upon further investigation, we determined that these reaction conditions denatured the protein. Despite being reported as a reaction under native conditions, protein denaturation is likely protein dependent. Fortunately, the refolding conditions for CREB are known and after refolding, all CREB variants displayed equal DNA binding affinities (Figure 6.7). These findings re-confirm that *O*-GlcNAcylation must disrupt the interaction of a cofactor protein opposed to directly affecting DNA binding.

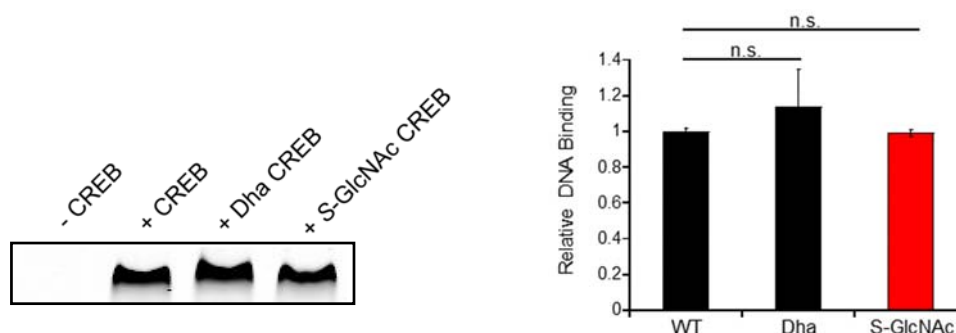


Figure 6.7. CREB DNA binding was unaffected by S40C modification as detected by electrophoretic mobility shift assay (EMSA). Both Dha-modified and S-GlcNAc-modified CREB did not show any significant change in DNA binding as compared to unmodified CREB indicating that modification at this position does not affect the DNA binding domain reconfirming findings from previous cellular studies.

Conclusion

In conclusion, here we have generated CREB site-specifically glycosylated at Ser40 with an *S*-GlcNAc analog and demonstrated the potential of using the proposed strategy for the synthesis of *O*-GlcNAcylated proteins. Chemical glycosylation was achieved through the unique conversion of a cysteine residue to Dha, followed by the nucleophilic addition of a thiol-containing GlcNAc analog. Reaction optimization and progress were monitored by LC-MS and indicated the complete conversion to the glycosylated product. Finally, the generation of functional protein was verified by the recapitulation of the DNA binding effects seen in a native cellular context.

The generation of homogeneously glycosylated CREB will allow for future experiments to further understand the structural and functional implications of this modification. Access to the glycosylated form will allow for functional studies into the

interplay between glycosylation and phosphorylation at the molecular level. The results should provide insights into potential combinatorial effects within the same protein molecule – a phenomenon that cannot be readily studied in cells. Moreover, with this method, we can also begin to dissect the functional roles of the six different glycosylation sites within CREB by generating each glycoprotein and evaluating their effects on protein interactions and transcriptional activity.

Studies have suggested long-range interactions between the Q2 and KID domains of CREB.²¹ Moreover, we found that glycosylation at Ser40 disrupts CREB's interaction with CRTC. Coupled with the observation that CRTC interacts primarily with the bZIP domain of CREB, this is indicative of potential long-range interactions. These findings, combined with the potential of *O*-GlcNAc modifications to induce structural change, make *O*-GlcNAc induced structural change in CREB one potential mechanism. Potential future work on elucidating the structural implications would include protein crystallization of the glycosylated protein or NMR studies of specific CREB domains.

Experimental Methods

Mutagenesis. Protocols were followed directly from Agilent QuikChange Lightning Mutagenesis Kit. Rat CREB in a pET23b vector was donated by Richard Goodman (OHSU). In short, primers were designed via Agilent website for each mutation separately. The provided reaction buffer and reagents were incubated with primers (IDT) and template vector and subjected to the recommended PCR conditions. The PCR product was digested with the provided DpnI and transformed into XL10-Gold ultracompetent cells.

CREB Bacterial Expression and Purification. Rat CREB was expressed as construct containing an N-terminally His₆-tag and TEV cleavage site in a pET23b vector. Expression was carried out in Rosetta 2 (DE3) pLysS cells. Transformed cells were grown at 37°C in 1L LB medium supplemented with 100 ug/mL ampicillin and 34 ug/mL chloramphenicol. At OD₆₀₀ = 0.8, cells were induced with the final concentration of 0.4 mM IPTG and left to shake an additional 3 hours at 37°C. Cells were pelleted at 4000 rpm at 4°C for 30 min and stored at -80°C until purification. The cells were resuspended in resuspension buffer (50 mM Tris pH 7.9, 100 mM KCl, 12.5 mM MgCl₂) and final concentrations of 25 ug/mL RNase, 2 mM dithiothreitol (DTT), 1x EDTA-free protease inhibitor cocktail, and 1 mM PMSF. The mixture was aliquoted into 1.5 mL centrifuge tubes and heated on a heating block at 65°C for 15 mins. Immediately after, the aliquots were put on ice for 10 mins and 2 mM DTT, 1x EDTA-free protease inhibitor cocktail (Roche), and 1 mM PMSF were added again. Cells were spun down at 13,000 rpm for 30 mins at 4°C. The soluble fraction was then added to 3 mL of pre-equilibrated Ni-NTA agarose and incubated for 1 hour at 4°C with end to end rotation. The resin was washed with 90 mL of washing buffer (10 mM Tris, 100 mM NaH₂PO₄, 8 M urea, pH 6.3) at RT. Protein was eluted with elution buffer (10 mM Tris, 100 mM NaH₂PO₄, 8 M urea, pH 5.9). The protein was then buffer exchanged into 100 mM Tris pH 8.0 with 30 kDa MWCO spin filters and incubated with His₆-TEV protease (0.5 mg/mL) in a 1:50 TEV:protein ratio for 1 h at RT. The mixture was then flowed over pre-equilibrated Ni-NTA agarose and the flowthrough was collected. The purified protein was then refolded by dialysis 3 times in 4 L of dialysis buffer (50 mM NaH₂PO₄ pH 8, 100 mM NaCl, 1 mM MgCl₂, 2 mM DTT).

Electrophoretic Mobility Shift Assay (EMSA). In a 20 μ L reaction, 1 μ g of CREB was incubated in a buffer containing a 20x EMSA buffer (200 mM Tris pH 7.5, 3 M KCl, 200 mM DTT, 5 mM EDTA), 0.1 μ g/ μ L poly(dIdC), 2% glycerol, and 1 μ L of CREB IRDye 700 (LICOR) at RT for 30 mins. The reaction was then separated on a 12% polyacrylamide gel in TAE buffer (40 mM Tris, 2.5 mM EDTA, pH to 7.8 with acetic acid) for 2.5 hours. Gels were imaged on an Odyssey Imager (LICOR).

***In Vitro* Transcription Assay.** Assays were performed using HeLaScribe Nuclear Extract (Promega) according to the manufacturer's protocol with minor modifications. Linearized CRE template DNA was prepared by digesting the plasmid pCRE-Luc (Stratagene) with NdeI and EcoRV. The resulting 2.1 kb fragment containing the CRE enhancer and partial luciferase ORF (~1.4 kb) was gel purified. Recombinantly expressed CREB (0.225 μ g) was combined with the CRE template DNA (100 ng), 0.4 mM rATP, 0.4 mM rUTP, 0.4 mM rCTP, 0.016 mM rGTP and 10 μ Ci [α - 32 P]rGTP (3000 Ci/mmol) in 7.3 mM HEPES pH 7.9, 6.1 mM MgCl₂, 37 mM KCl, 0.07 mM EDTA, 0.2 mM DTT, 5.5% glycerol. Transcription reactions were initiated by the addition of HeLaScribe nuclear extract (8 U). Run-off RNA transcripts were resolved on 7 M urea 6% TBE-PAGE gels and visualized by autoradiography using the Typhoon FLA 7000.

Bioconjugation Reaction of Cys \rightarrow Dha \rightarrow S-GlcNAc. Recombinantly-expressed, mutant CREB was buffer exchanged and concentrated to 2 mg/mL using 3 kDa MWCO spin filters (Millipore) into a 50 mM NaH₂PO₄ pH 8.0 at 4 $^{\circ}$ C. MSH (14 mg/mL) or

DBHDA (132 mg/mL) in DMF was added at the appropriate equivalence to the protein solution and quickly mixed by pipetting. MSH reactions were carried out at 4 °C for 20 min with end-over-end rotation. DBHDA reactions were first incubated at RT for 30 min then moved to 37 °C for 90 min with end-over-end rotation. After the indicated times, an aliquot was removed for later analysis by LC-MS to confirm the conversion to Dha. Solid 1-thio-2-acetamido-2-deoxy- β -D-glucopyranose was then added to the solution (5000 eq.) and the reaction was incubated at 37 °C for 2 h with end-over-end rotation. After 2 h, the reaction was quenched by dilution with 10 mM Tris, 100 mM NaH₂PO₄, 8 M urea, pH 6.3 and refolded by dialysis 3 times in 4 L of dialysis buffer (50 mM NaH₂PO₄ pH 8, 100 mM NaCl, 1 mM MgCl₂, 2 mM DTT) with 12 kDa MWCO dialysis tubing. The final protein was then analyzed by LC-MS and SDS-PAGE for reaction completion and purity.

Intact Protein Mass Spectrometry. Liquid chromatography-mass spectrometry (LC-MS) was performed on an Agilent 1100 Series HPLC coupled to an Agilent 1100 Series MSD connected to an Agilent Zorbax 300SB-C3 (Narrowbore 2.1x150mm, 5 μ m) column. The sample was buffer exchanged into 2% formic acid with 3 kDa MWCO spin filters and injected and subjected to 25 min of 0.2% formic acid in water to desalt. Then a linear gradient was applied for 25 min to 95% acetonitrile with 0.2% formic acid and 10% methanol. All of this was performed at a flow rate of 0.2 mL/min. After the run, the column was equilibrated for 10 min with solvent A. The electrospray source had a capillary voltage of 4.0 kV. Nitrogen was used as the nebulizer and desolvation gas with a flow rate of 3 L/min. Protein will typically elute between 38 and 46 min. ChemStation from Agilent was

used to analyze and deconvolute mass ion data. The system was calibrated with Agilent ESI Tuning Mix and accuracy was confirmed with a standard sample of apomyoglobin. Protein purity was determined by the ratio of product ion abundance versus the abundance of the total protein ions found with a 5% protein abundance cutoff.

Peptide Identification by LC-MS/MS. A directed mass spectrometric search was used to determine the positions of Dha and S-GlcNAc labeled peptides. Semi-synthetically prepared CREB was analyzed by SDS-PAGE (4-12% Bis-Tris gels) and the resulting bands were cut out. In-gel digestion of gel bands with chymotrypsin was performed as previously described.²² To help identify peptides corresponding to Dha and S-GlcNAc modifications, an m/z inclusion list for expected peptide ions was generated. LC-MS/MS experiments were performed on an LTQ-FT Ultra mass spectrometer (Thermo Fisher Scientific) equipped with a nanoelectrospray ion source (Thermo Fisher Scientific) connected to an EASY-nLC II (Thermo Fisher Scientific). MS/MS spectra were collected in CID mode and target MS/MS were searched using the ROCCIT MS/MS Search Engine.

Wheat Germ Agglutinin (WGA) Affinity Purification. Protocol was followed from *Curr. Protoc. Protein Sci.* 2011, Chapter 12, Unit 12.8.²³ In short, semi-synthetically prepared CREB was incubated end-over-end with WGA-agarose (Vector Labs) pre-equilibrated with 10 mM phosphate buffer pH 7.8, 150 mM NaCl (Washing Buffer) for 30 min at 4 °C. After 30 min, the beads were applied to a column and washed with 4 column volumes of Washing Buffer and 4 column volumes of 10 mM phosphate buffer pH 7.8,

150 mM NaCl, 1 M D-galactose (Gal Buffer). CREB was then eluted with 7 column volumes of 10 mM phosphate buffer pH 7.8, 150 mM NaCl, 1 M D-*N*-acetylglucosamine (GlcNAc Elution Buffer). Fractions were collected and analyzed by SDS-PAGE.

General Methods for Chemical Synthesis. All reactions were carried out under an argon atmosphere unless otherwise specified. Reagents were purchased from Sigma Aldrich and used without purification unless otherwise specified. ^1H NMR spectra were recorded with a Varian Inova 500 (500 MHz) and are reported relative to residual solvent peaks. Data for ^1H NMR are reported as follows: chemical shift (δ ppm) (multiplicity, coupling constant (Hz), integration, assignment). Multiplicities are reported as follows: s = singlet, d = doublet, t = triplet, m = multiplet, br = broad. Data for ^{13}C NMR are reported in terms of chemical shifts (δ ppm).

Synthesis of ethyl-O-(mesitylenesulfonyl)acetohydroxamate. 3.5 g (34.0 mmol) of ethyl *N*-hydroxyacetimidate (Alfa Aesar) was dissolved into 9 mL of DMF. 4.5 mL of triethylamine (Alfa Aesar) was added and the solution was stirred at 0°C for 15 mins. 7.5 g (34.3 mmol) of 2-mesitylenesulfonyl chloride (TCI America) was added in two portions and stirred vigorously for 30 mins at 0°C. Mixture took on an orange sticky consistency. Mixture was diluted with DCM and extracted with H_2O to remove DMF. The organic layer was dried with MgSO_4 , filtered, and solvent removed under reduced vacuum. The final product was used without further purification as a white solid in 83% yield. ^1H NMR (500 MHz, CDCl_3) 1.20 (3H, t, $J = 7.1$ Hz, CH_2CH_3), 2.04 (3H, s, CH_3), 2.31 (3H, s, CH_3Ar), 2.64 (6H, s, 2 x CH_3Ar), 3.90 (2H, q, $J = 7.1$ Hz, CH_2CH_3), 6.97 (2H, s, Ar-H).

Synthesis of *O*-mesitylenesulfonylhydroxylamine (MSH). Caution, MSH can detonate if crystallized improperly. 7.1 g (24.8 mmol) of 1 was added to 6.4 mL of p-dioxane and stirred at 0°C for 20 min. 2.9 mL of 70% perchloric acid (Sigma) was added slowly dropwise via pipette over 2 min and maintaining the temperature under 10°C. After 10 min of stirring at 0°C, the mixture solidified. The mixture was then transferred to a 1 L Erlenmeyer flask with 320 mL of ice water and rinsed with water and ethyl ether. The contents were then extracted with more ethyl ether. The organic layer was neutralized with potassium carbonate for 5 min and filtered. The filtrate was then concentrated to less than 160 mL and poured into 240 mL of ice cold hexanes. Crystallization was allowed to proceed overnight at 4°C. Small white needles were crystallized, filtered, and transferred to a plastic Falcon tube. The compound was further dried under vacuum and stored at -20°C at 41% yield. ¹H NMR (500 MHz, CDCl₃) 2.33 (3H, s, CH₃Ar), 2.65 (6H, s, 2 x CH₃Ar), 3.85 (2H, br, NH₂), 7.00 (2H, s, Ar-H). MS-TOF: [M+H]⁺ 216.2515.

Synthesis of 2,5-dibromohexanediamide (DBHDA). Adipic acid (5.01 g, 34.3 mmol) was added to a 100 mL round bottom flask and suspended in thionyl chloride (15 mL, 207 mmol). The mixture was refluxed for 1.5 h while stirring open to air at 80 °C. After 1.5 h, the reaction was cooled to RT. CCl₄ (20 mL) was added to the reaction followed by *N*-bromosuccinimide (14.62 g, 82.2 mmol). The reaction was stirred vigorously and 2 drops of HBr was added by pipette. The reaction was refluxed again open to air at 85 °C for 2 h. After 2 h, the reaction was cooled to RT and then 0 °C. The mixture was stirred at 0 °C to ensure the precipitation of succinimide. The solid was removed by filtration and rinsed

with Et₂O. The filtrate was concentrated *in vacuo* to yield a thick black liquid. In a separate 100 mL round bottom flask, 40 mL of NH₄OH (28-30% aqueous) was cooled to 0 °C. The crude acid chloride was added dropwise to the flask over 20 minutes. After the addition is complete, stirring continued for 1 h. The green solid was vacuum filtered and collected. The product was purified by triturating in 20 mL of 1:1 MeOH/H₂O and heated to 60 °C. After cooling to RT, the white solid was collected by vacuum filtration, rinsed with MeOH, and dried under vacuum (5.21 g, 50%). ¹H NMR (500 MHz, DMSO-d₆) 1.78-2.07 (4H, m, CH₂CH₂), 4.29-4.34 (2H, m, 2 x CHBr), 7.31, (2H, s, NH₂), 7.69 (2H, s, NH₂).

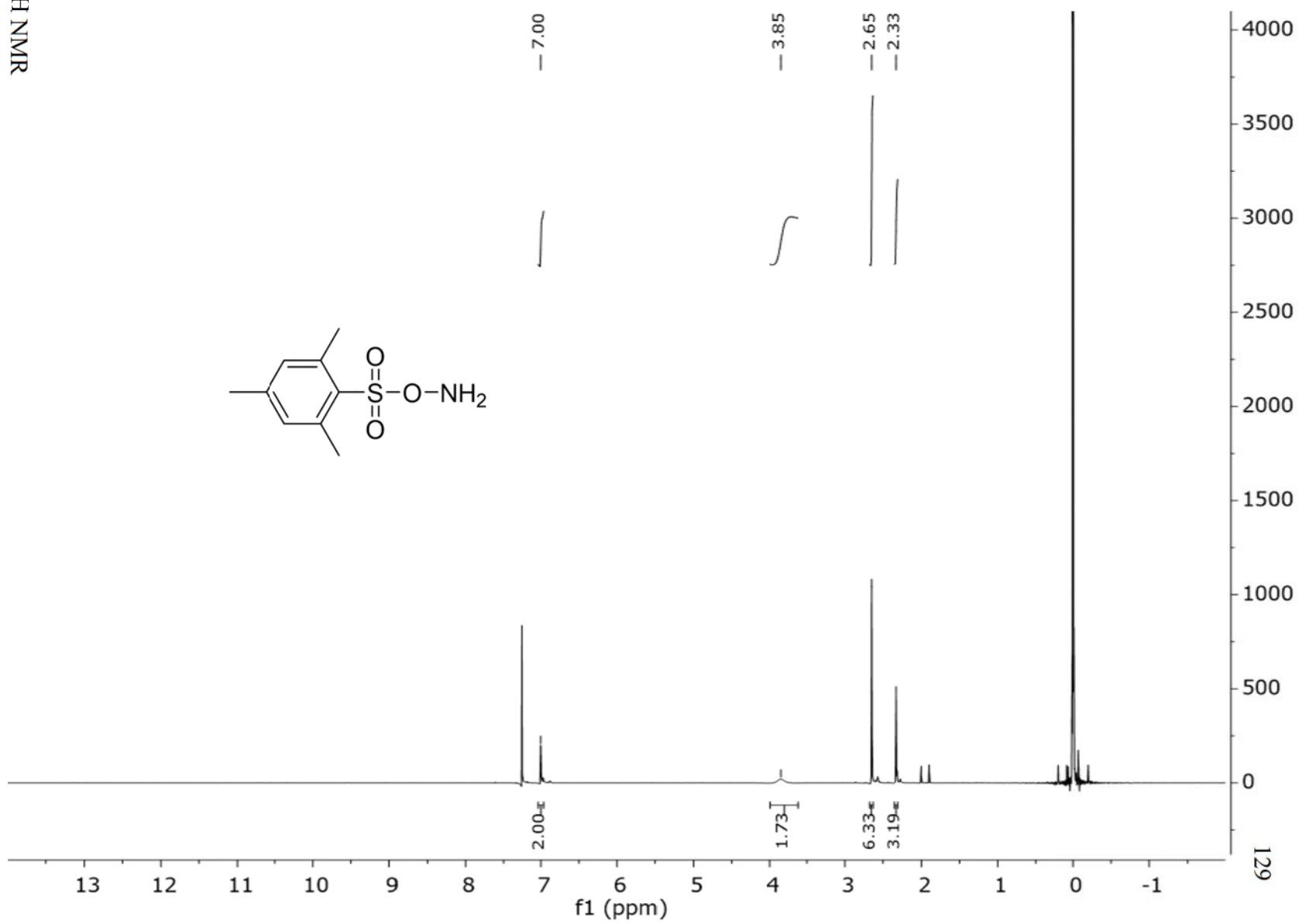
Synthesis of (3,4,6-tri-*O*-acetyl-2-acetamido-2-deoxy-β-Dglucopyranosyl)-1-isothiuronium chloride. 4.0 g (TCI America, 10.9 mmol) of 3,4,6-tri-*O*-acetyl-2-acetamido-2-deoxy-α-D-glucopyranosyl chloride and 1.6 g (21.0 mmol) thiourea were dissolved into 34 mL of anhydrous acetone. The mixture was refluxed under argon for 2 hours at 60°C. The precipitate that built up was removed and the mixture was returned to reflux until there was no more precipitation. The white solid was then rinsed with acetone and hexanes, filtered, and dried under vacuum for an 89% yield. ¹H NMR (500 MHz, DMSO-d₆) 1.81 (3H, s, NHCOCH₃), 1.95 (3H, s, CH₃), 1.99 (3H, s, CH₃), 2.02 (3H, s, CH₃), 4.06 (1H, dd, *J*_{5,6} = 2.2 Hz, *J*_{6,6'} = 10.0 Hz, H-6), 4.17 (1H, dd, *J*_{5,6'} = 5.0 Hz, *J*_{6,6'} = 12.3 Hz, H-6'), 4.26 (1H, ddd, *J*_{4,5} = 10.2 Hz, *J*_{5,6} = 2.3 Hz, *J*_{5,6'} = 4.7 Hz, H-5), 4.94 (1H, at, *J* 9.8 Hz, H-4), 5.13 (1H, at, *J* 9.7 Hz, H-3), 5.63 (1H, d, *J*_{1,2} 10.3 Hz, H-1), 8.40 (1H, d, *J* 9.3 Hz, NHAc), 9.20 (4H, brs, 2 x NH₂).

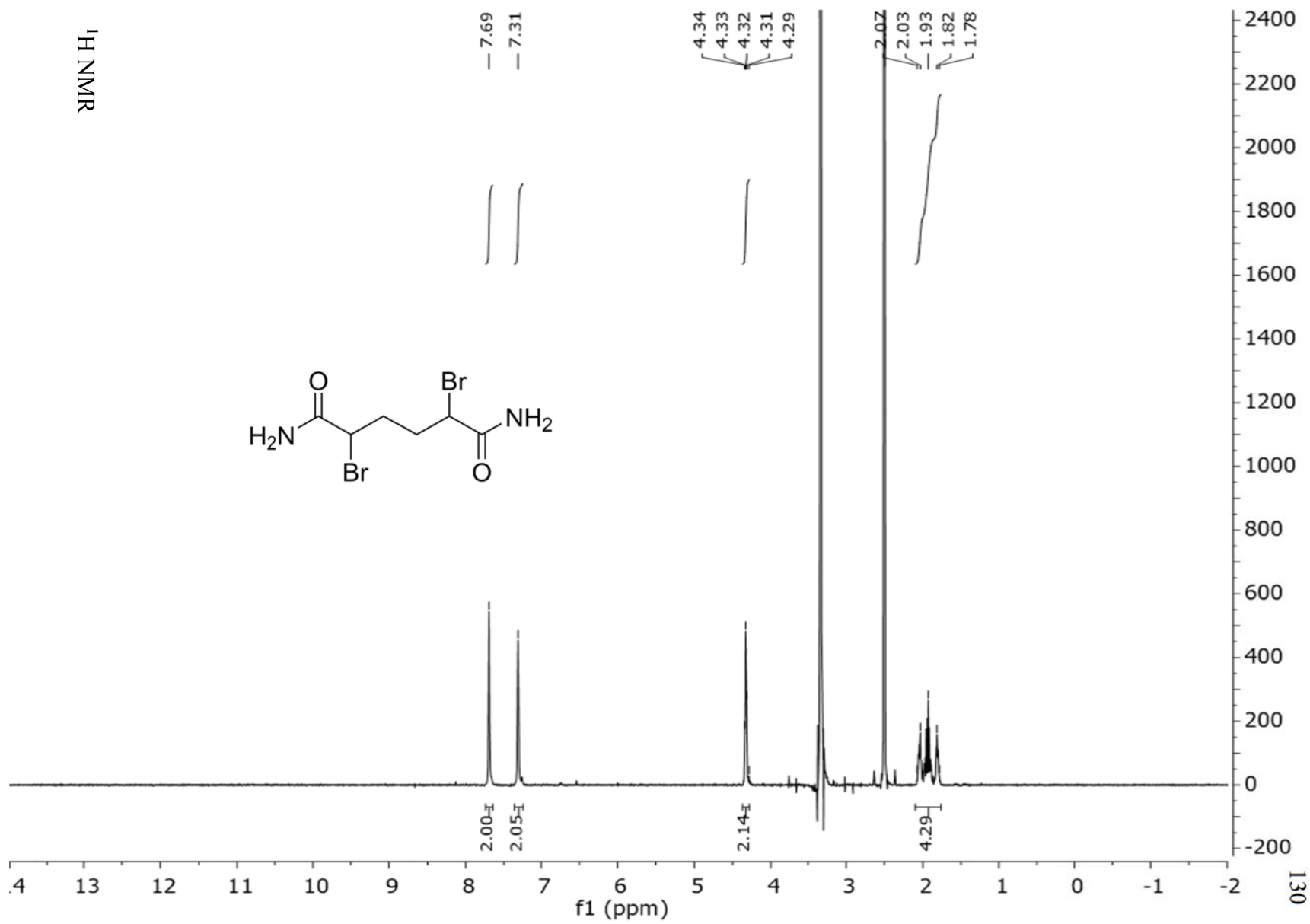
Synthesis of 1-thio-3,4,6-tri-*O*-acetyl-2-acetamido-2-deoxy-β-D-glucopyranose. 4.2 g (9.51 mmol) of (3,4,6-tri-*O*-acetyl-2-acetamido-2-deoxy-β-D-glucopyranosyl)-1-

isothiuronium chloride was dissolved in DCM and H₂O by stirring at RT. 2.2 g (11.6 mmol) of sodium metabisulfite was then added to the mixture and refluxed at 46°C for 2 hours. The reaction was left to cool to RT after 2 hours. The mixture was then extracted twice with DCM. All organic layers were combined and washed with H₂O and brine and dried over MgSO₄. The solvent was removed by under reduced vacuum and recrystallized in hot EtOAc and hexanes. The white solid was filtered and dried under vacuum to a yield of 79% yield. ¹H NMR (500 MHz, CDCl₃) 1.99 (3H, s, CH₃), 2.03 (3H, s, CH₃), 2.05 (3H, s, CH₃), 2.10 (3H, s, CH₃), 2.57 (1H, d, $J_{1,SH} = 9.4$ Hz, SH), 3.68 (1H, ddd, $J_{4,5} = 9.8$ Hz, $J_{5,6} = 4.8$ Hz, $J_{5,6'} = 2.2$ Hz, H-5), 4.10-4.16 (2H, m, H-2, H-3), 4.24 (1H, dd, $J_{5,6} = 4.8$ Hz, $J_{6,6'} = 12.5$ Hz, H-6), 4.56 (1H, t, $J = 10.3$ Hz, H-1), 5.05-5.15 (2H, m, H-4, H-6'), 5.55 (1H, d, $J = 9.4$ Hz, NH).

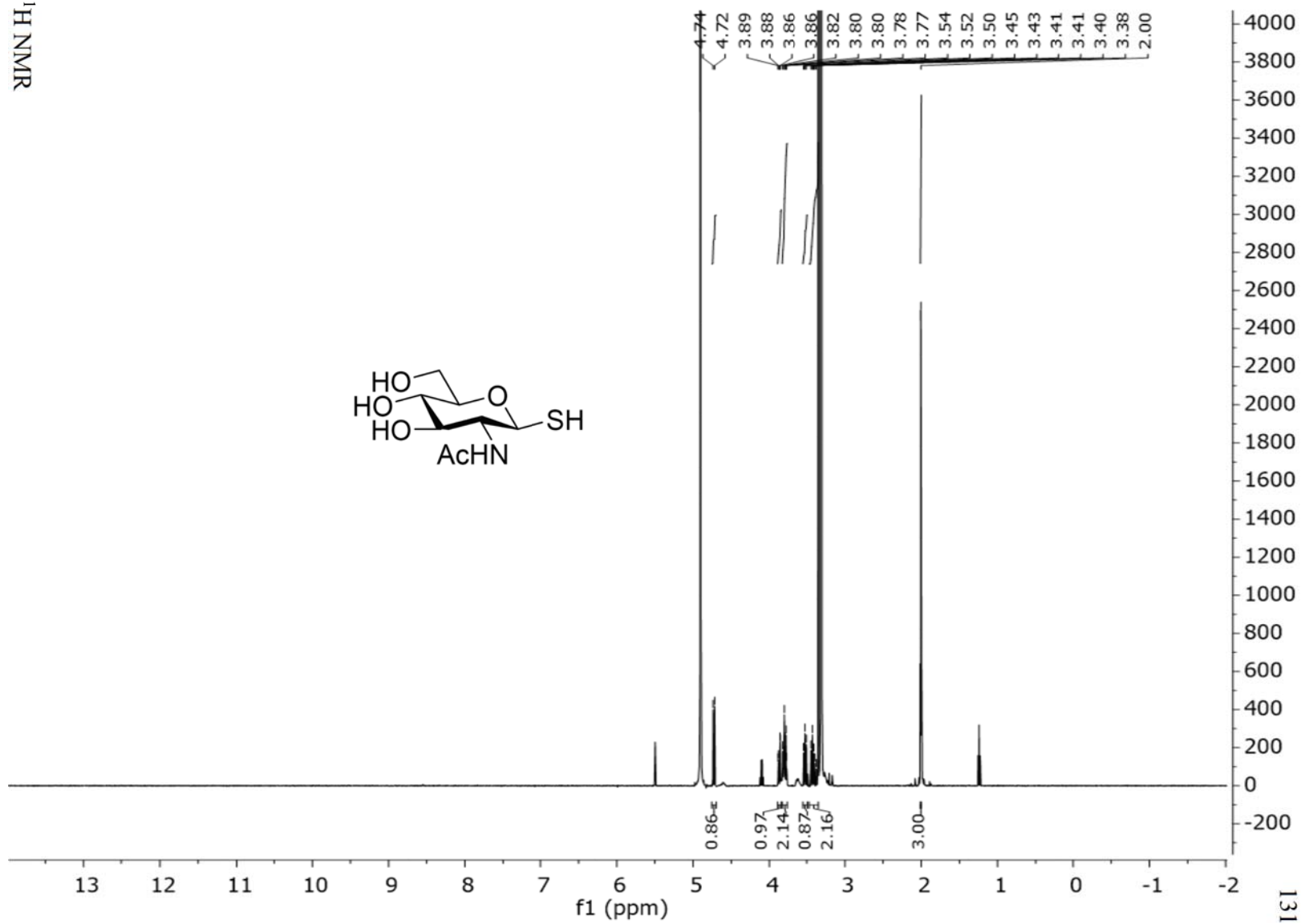
Synthesis of 1-thio-2-acetamido-2-deoxy- β -D-glucopyranose. 2.3 g (6.33 mmol) of 3,4,6-tri-*O*-acetyl-2-acetamido-2-deoxy- β -D-glucopyranosyl thiol was stirred into MeOH at RT. 1.1 equivalence of NaOMe (Sigma) was added and the mixture was stirred and monitored for 2.5 hours by TLC (40% MeOH/DCM). The reaction was neutralized with Dowex-50 ion exchange resin (Alfa Aesar). The mixture was filtered and concentrated under reduced vacuum. The product was recrystallized in MeOH/EtOAc to a sticky oil in 40% yield. ¹H NMR (500 MHz, MeOH), 2.00 (3H, s, CH₃), 3.38-3.45 (2H, m, H-4, H-5), 3.52 (1H, t, $J = 8.9$ Hz, H-3), 3.77-3.82 (2H, m, H-2, H-6), 3.89 (1H, dd, $J_{5,6'} = 2.2$ Hz, $J_{6,6'} = 11.9$ Hz, H-6'), 4.74 (1H, d, $J_{1,2} = 10.4$ Hz, H-1). MS-TOF: [M-H]⁻ 236.0628.

¹H NMR





¹H NMR



References

- (1) Slawson, C.; Hart, G. W. *Nat. Rev. Cancer* **2011**, *11*, 678.
- (2) Yi, W.; Clark, P. M.; Mason, D. E.; Keenan, M. C.; Hill, C.; Goddard, W. A.; Peters, E. C.; Driggers, E. M.; Hsieh-Wilson, L. C. *Science* **2012**, *337*, 975.
- (3) Ma, Z. Y.; Vosseller, K. *Amino Acids* **2013**, *45*, 719.
- (4) Slawson, C.; Copeland, R. J.; Hart, G. W. *Trends Biochem. Sci.* **2010**, *35*, 547.
- (5) Banerjee, P. S.; Ma, J. F.; Hart, G. W. *P. Natl. Acad. Sci. U.S.A.* **2015**, *112*, 6050.
- (6) Wells, L.; Vosseller, K.; Hart, G. W. *Science* **2001**, *291*, 2376.
- (7) Rexach, J. E.; Clark, P. M.; Hsieh-Wilson, L. C. *Nat. Chem. Biol.* **2008**, *4*, 97.
- (8) Yuzwa, S. A.; Shan, X. Y.; Macauley, M. S.; Clark, T.; Skorobogatko, Y.; Vosseller, K.; Vocadlo, D. J. *Nat. Chem. Biol.* **2012**, *8*, 393.
- (9) Lazarus, B. D.; Love, D. C.; Hanover, J. A. *Int. J. Biochem. Cell B.* **2009**, *41*, 2134.
- (10) Lamarre-Vincent, N.; Hsieh-Wilson, L. C. *J. Am. Chem. Soc.* **2003**, *125*, 6612.
- (11) Rexach, J. E.; Clark, P. M.; Mason, D. E.; Neve, R. L.; Peters, E. C.; Hsieh-Wilson, L. C. *Nat. Chem. Biol.* **2012**, *8*, 253.
- (12) Bernardes, G. J. L.; Chalker, J. M.; Errey, J. C.; Davis, B. G. *J. Am. Chem. Soc.* **2008**, *130*, 5052.
- (13) Gamblin, D. P.; Garnier, P.; van Kasteren, S.; Oldham, N. J.; Fairbanks, A. J.; Davis, B. G. *Angew. Chem. Int. Edit.* **2004**, *43*, 828.
- (14) Lercher, L.; Raj, R.; Patel, N. A.; Price, J.; Mohammed, S.; Robinson, C. V.; Schofield, C. J.; Davis, B. G. *Nat. Commun.* **2015**, *6*.
- (15) Schumacher, M. A.; Goodman, R. H.; Brennan, R. G. *J. Biol. Chem.* **2000**, *275*, 35242.
- (16) Lopez, D. I.; Mick, J. E.; Nyborg, J. K. *Protein Expres. Purif.* **2007**, *55*, 406.
- (17) Santiagorivera, Z. I.; Williams, J. S.; Gorenstein, D. G.; Andrisani, O. M. *Protein Sci.* **1993**, *2*, 1461.

- (18) Luo, Q.; Viste, K.; Urdy-Zaa, J. C.; Kumar, G. S.; Tsai, W. W.; Talai, A.; Mayo, K. E.; Montminy, M.; Radhakrishnan, I. *P. Natl. Acad. Sci. U.S.A.* **2012**, *109*, 20865.
- (19) Chalker, J. M.; Lercher, L.; Rose, N. R.; Schofield, C. J.; Davis, B. G. *Angew. Chem. Int. Edit.* **2012**, *51*, 1835.
- (20) Fernandez-Gonzalez, M.; Boutureira, O.; Bernardes, G. J. L.; Chalker, J. M.; Young, M. A.; Errey, J. C.; Davis, B. G. *Chem. Sci.* **2010**, *1*, 709.
- (21) Asahara, H.; Santoso, B.; Guzman, E.; Du, K. Y.; Cole, P. A.; Davidson, I.; Montminy, M. *Mol. Cell Biol.* **2001**, *21*, 7892.
- (22) Clark, P. M.; Dweck, J. F.; Mason, D. E.; Hart, C. R.; Buck, S. B.; Peters, E. C.; Agnew, B. J.; Hsieh-Wilson, L. C. *J. Am. Chem. Soc.* **2008**, *130*, 11576.
- (23) Zachara, N. E.; Vosseller, K.; Hart, G. W. *Curr. Protoc. Protein Sci.* **2011**, *Chapter 12*, Unit 12.8.

Appendix A

SYNTHESIS OF URIDINE DIPHOSPHATE 6-AZIDO D-GALACTOSE

A p p e n d i x A

SYNTHESIS OF URIDINE DIPHOSPHATE 6-AZIDO D-GALACTOSE

Experimental Methods

Compound was prepared as previously reported in *Tetrahedron Lett.* **2008**, 49, 2294 with some modifications.

General Methods. All reactions were carried out under an argon atmosphere unless otherwise specified. Reagents were purchased from Sigma Aldrich and used without purification unless otherwise specified. ^1H NMR spectra were recorded with a Varian Inova 500 (500 MHz) and are reported relative to residual solvent peaks. ^{31}P NMR spectra was recorded with a Varian Inova 300 (300 MHz). Data for ^1H NMR are reported as follows: chemical shift (δ ppm) (multiplicity, coupling constant (Hz), integration, assignment). Multiplicities are reported as follows: s = singlet, d = doublet, t = triplet, m = multiplet, br = broad. Data for ^{13}C NMR are reported in terms of chemical shifts (δ ppm).

Synthesis of 1,2,3,4-di-*O*-isopropylidene-6-azido-6-deoxy- α -D-galactopyranose

1,2:3,4-Di-*O*-isopropylidene- α -D-galactopyranose (4.5 g, 17.3 mmol) and 4-dimethylaminopridine (DMAP, 4.22 g, 2 eq) was dissolved into DCM (82 mL) and stirred at 0 °C under argon. Methanesulfonyl chloride (MsCl, 2.0 mL 1.5 eq) was added dropwise.

The reaction was allowed to stir for 15 min at 0 °C, then warmed to RT and stirred for 1 h. After 1 h, the reaction was diluted with DCM (82 mL), filtered through celite, and washed with DCM (82 mL). The filtrate was concentrated and dissolved into DMF (82 mL). Tetrabutylammonium bromide (5.56 g, 1 eq) and sodium azide (10.2 g, 9 eq) were successively added to the reaction mixture and was warmed to 110 °C under argon. After 5 days, the reaction mixture was cooled to room temperature and concentrated *in vacuo*. Saturated NaHCO₃ (164 mL) and DCM (164 mL) were added to the crude compound and aqueous phase was extracted with DCM. The organic phase was then washed with water (164 mL), brine (328 mL), dried over MgSO₄, and concentrated *in vacuo*. Purification of the residue by flash chromatography (9:1 Hex/EtOAc) afforded the compound (4.04 g, 82%). ¹H NMR (500 MHz, CDCl₃) δ 1.3 (6H, s, CCH₃), 1.42 (3H, s, CCH₃), 1.51 (3H, s, CCH₃), 3.29 (1H, dd, H-6'), 3.44 (1H, dd, *J*_{6,6'} = 12.7 Hz, H-6), 3.88 (1H, ddd, *J*_{5,6} = 7.7 Hz, *J*_{5,6'} = 5.4 Hz, H-5), 4.14 (1H, dd, *J*_{4,5} = 1.9 Hz, H-4), 4.57 (1H, dd, *J*_{2,3} = 2.5 Hz, *J*_{3,4} = 2.5 Hz, H-3), 5.50 (1H, d, *J*_{1,2} = 5.0 Hz, H-1).

Synthesis of 6-azido-6-deoxy-D-galactopyranose

The previously made 1,2,3,4-di-*O*-isopropylidene-6-azido-6-deoxy- α -D-galactopyranose (4.04 g, 14.1 mmol) was dissolved into hydrochloric acid (HCl, 0.6 M, 551 mL) and warmed to 80 °C. After 18 h, the pH was adjusted to 7 with solid sodium bicarbonate (0.21 g). The reaction mixture was lyophilized and purified by flash chromatography (20% MeOH/DCM) to yield a white solid (2.76 g, 95%). ¹H NMR (500 MHz, D₂O) δ 3.33-3.65 (m, 3.4H, H-3 β , H-6 β , H-6 β ', H-6 α , H-6' α , H-2 β), 3.75-3.83 (m, 1.3H, H-5 β , H-3 α , H-2 α),

3.87 (m, 0.7H, H-4 β), 3.92 (m, 0.3H, H-4 α), 4.15 (m, 0.3H, H-5 α), 4.56 (d, 0.7H, $J_{1-2} = 8.2$ Hz, H-1 β), 5.21 (d, 0.3H, $J_{1,2} = 3.6$ Hz, H-1 α).

Synthesis of 1,2,3,4-tetra-*O*-acetyl-6-azido-6-deoxy- α -D-galactopyranose

Acetic anhydride (14.7 mL) was added to a round bottom flask containing the previously synthesized 6-azido-6-deoxy-D-galactopyranose (2.76 g, 13.5 mmol) at 0 °C under argon. Boron trifluoroetherate (3.5 mL, 1.8 eq) was added dropwise to the suspension. After 5 h at RT, the reaction mixture was diluted with DCM (200 mL) and extracted with sat. NaHCO₃ (3 x 200 mL) and brine (100 mL). The organic layer was dried with MgSO₄, concentrated *in vacuo*, and purified by flash chromatography (30% EtOAc/Hex) to yield a white solid (4.29 g, 80%). ¹H-NMR (500 MHz, CDCl₃) δ 2.01 (3H, s, CH₃COO), 2.03 (3H, s, CH₃COO), 2.17 (3H, s, CH₃COO), 2.18 (3H, s, CH₃COO), 3.19 (1H, dd, H-6'), 3.42 (1H, dd, $J_{6,6'} = 12.4$ Hz, H-6), 4.23 (1H, br dd, $J_{5,6} = 7.4$ Hz, $J_{5,6'} = 5.4$ Hz, H-5), 5.33-5.34 (2H, m, H-2, H-3), 5.48 (1H, d, $J < 1$ Hz, H-4), 6.40 (1H, s, H-1).

Synthesis of 2,3,4-tri-*O*-acetyl-6-azido-6-deoxy- α -D-galactopyranose

A solution of 1,2,3,4-tetra-*O*-acetyl-6-azido-6-deoxy- α -D-galactopyranose (3.89 g, 10.4 mmol) in DMF (80 mL) and *N,N*-diisopropylethylamine (DIPEA, 15.3 mL, 8.6 eq) was stirred with ammonium acetate crystals (3.24 g, 4 eq) at RT for 2 days. After 2 days, the reaction mixture was decanted from the ammonium acetate crystals into a separatory funnel and diluted with DCM (260 mL). The mixture was then extracted with sat. NaHCO₃ (170

mL) and water (80 mL), dried with MgSO_4 , concentrated *in vacuo*, and purified by flash chromatography (1:1 EtOAc/Hex) to yield a clear oil (2.76 g, 80%). $^1\text{H-NMR}$ (500 MHz, CDCl_3) δ 1.95 (3H, s, $\text{CH}_3\text{COO}_\beta$), 2.06 (3H, s, $\text{CH}_3\text{COO}_\alpha$), 2.12 (3H, s, CH_3COO), 3.20 (0.83H, dd, H-6' $_\alpha$), 3.24 (0.17H, dd, H-6' $_\beta$), 3.42 (0.83H, dd, $J_{6,6'} = 12.8$ Hz, H-6 $_\alpha$), 3.51 (0.17H, dd, $J_{5,6} = 8.0$ Hz, $J_{6,6'} = 12.8$ Hz, H-6 $_\beta$), 3.83 (ddd, 0.17H, $J_{4,5} = 0.7$ Hz, $J_{5,6} = 7.6$ Hz, $J_{5,6'} = 5.0$ Hz, H-5 $_\beta$), 4.08 (0.17H, d, OH $_\beta$), 4.27 (0.83H, br, OH $_\alpha$), 4.37 (0.83H, dd, $J_{5,6} = 4.4$ Hz, $J_{5,6'} = 8.3$ Hz, H-5 $_\alpha$), 4.72 (0.17H, t, $J_{1,\text{OH}} = J_{1,2} = 6.8$ Hz, H-1 $_\beta$), 4.98-5.06 (0.34H, m, H- β , H-2 $_\beta$), 5.12 (0.83H, dd, $J_{2,3} = 10.0$ Hz, H-2 $_\alpha$), 5.32-5.38 (1H, m, H- β , H-3 $_\alpha$), 5.49 (0.83H, s, H-4 $_\alpha$).

Synthesis of 2,3,4-tri-*O*-acetyl-6-azido-6-deoxy- α -D-galactopyranosyl phosphite mono- Et_3N salt

The previously synthesized 2,3,4-tri-*O*-acetyl-6-azido-6-deoxy- α -D-galactopyranose (2.76 g, 8.3 mmol) was dissolved in THF (11 mL) and dioxane (22 mL) and stirred at 0 °C. Triethylamine (Et_3N , 1.77 mL, 1.55 eq) and a solution of 2-chloro-1,3,2-benzodioxaphosphorin-4-one (1.69 g, 1 eq) in THF (5.5 mL) was then added to the mixture and stirred for 35 min at 0 °C. The reaction was quenched with the addition of water (5 mL) and concentrated *in vacuo*. The concentrate was then redissolved into THF (110 mL) and the precipitant that formed was separated by filtration. After concentration of the filtrate, the crude compound was purified by column chromatography (gradient of 5% MeOH/DCM to 20% MeOH/DCM) to afford the product phosphite as a white solid (2.77 g, 84%). $^1\text{H-NMR}$ (500 MHz, CDCl_3) δ 1.32 (9H, t, $(\text{CH}_3\text{CH}_2)_3\text{NH}^+$), 1.96 (3H, s,

CH_3COO), 2.07 (3H, s, CH_3COO), 2.16 (3H, s, CH_3COO), 3.19 (6H, q, $J = 7.3$ Hz, $(CH_3CH_2)_3NH^+$), 3.28 (1H, dd, H-6'), 3.48 (1H, dd, $J_{6,6'} = 12.8$ Hz, H-6), 4.40 (1H, ddd, $J_{5,6} = 7.4$ Hz, $J_{5,6'} = 5.5$ Hz, H-5), 5.15 (1H, ddd, $J_{2,4} = 1.8$ Hz, H-2), 5.36 (1H, dd, $J_{2,3} = 10.8$ Hz, H-3), 5.46 (1H, br dd, $J_{3,4} = 3.3$ Hz, $J_{4,5} = 1.1$ Hz, H-4), 5.82 (1H, dd, $J_{1,2} = 3.4$ Hz, $J_{H,P} = 8.8$ Hz, H-1), 6.81 (1H, m, P-H).

Synthesis of 2,3,4-tri-*O*-acetyl-6-azido-6-deoxy- α -D-galactopyranosyl phosphate mono- Et_3N salt

The previously synthesized phosphite (2.77 g, 5.6 mmol) was dissolved into THF (100 mL) and stirred with cation exchange resin (Dowex 50W-X8 H^+ form) for 1 h at RT. The resin was filtered out and the filtrate was concentrated *in vacuo*. The resulting oil was then dissolved with THF (96 mL) and t-BuOOH (2.2 mL of 5.5 M solution in undecane, 2 eq) and iodine (19.3 mg, 0.01 eq) were added under argon. After stirring for 24 h at RT, the reaction mixture was neutralized with Et_3N (0.8 mL, 2.05 eq) and concentrated *in vacuo*. The concentrate was purified by a C18 reverse-phase column (water as eluent) to give the phosphate product as an oil (700 mg, 24%). 1H -NMR (500 MHz, CD_3OD) δ 1.32 (9H, t, $(CH_3CH_2)_3NH^+$), 1.97 (3H, s, CH_3COO), 2.06 (3H, s, CH_3COO), 2.16 (3H, s, CH_3COO), 3.17 (6H, q, $J = 7.2$ Hz, $(CH_3CH_2)_3NH^+$), 3.29 (1H, dd, H-6'), 3.50 (1H, dd, $J_{6,6'} = 12.5$ Hz, H-6), 4.41 (1H, ddd, $J_{5,6} = 6.2$ Hz, $J_{5,6'} = 7.1$ Hz, H-5), 5.11 (1H, ddd, H-2), 5.40 (1H, dd, $J_{2,3} = 10.9$ Hz, H-3), 5.47 (1H, dd, $J_{3,4} = 4.5$ Hz, $J_{4,5} < 1$ Hz, $J_{2,4} = 2.2$ Hz, H-4), 5.74 (1H, dd, $J_{1,2} = 2.8$ Hz, $J_{H,P} = 7.5$ Hz, H-1).

Synthesis of uridine 5'-(2,3,4-tri-*O*-acetyl-6-azido-6-deoxy- α -D -galactopyranosyl) diphosphate bistriethylammonium salt

4'-morpholine-*N,N'*-dicyclohexylcarboxamidinium (UMP-morpholidate, 1.02 g, 1.5 mmol, 2 eq) and tetrazole (3.45 mL, 0.45 M in acetonitrile, 2 eq) were added to a round bottom and dried three times with pyridine *in vacuo*. In another round bottom, the previously synthesized phosphate (380 mg, 0.74 mmol) was also dried three times with pyridine *in vacuo*. After both were dry, the UMP-morpholidate and tetrazole were dissolved in dry pyridine (70 mL) and added to the dried phosphate. The mixture was then dried two more times with pyridine *in vacuo*. Dry pyridine (30 mL) was then added after the materials were dry and the reaction was stirred in the dark for 2 days under argon. After 2 days, toluene was added and the mixture was concentrated *in vacuo*. After lyophilization, the product was purified on a semi-preparative HPLC C18-column (linear gradient for 10 min from 90:10 water/MeCN to 50:50 water/MeCN with detection at 254 nm) to afford the product (270 mg, 40%). ¹H-NMR (500 MHz, D₂O) δ 1.29 (18H, t, (CH₃CH₂)₃NH⁺), 1.99 (3H, s, CH₃COO), 2.12 (3H, s, CH₃COO), 2.19 (3H, s, CH₃COO), 3.53 (1H, dd, $J_{6-6'} = 12.9$ Hz, H-6_{gal}), 3.41 (q, 12H, $J = 7.3$ Hz, (CH₃CH₂)₃NH⁺), 4.16 (1H, ddd, $J = 3.3$ Hz, $J = 5.4$ Hz, $J_{5,5'} = 12.0$ Hz, H-5'_{rib}), 4.22-4.27 (2H, m, H-4_{rib}, H-5_{rib}), 4.31-4.36 (2H, m, H-2_{rib}, H-3_{rib}), 4.49 (1H, t, $J_{5,6,6'} = 6.7$ Hz, H-5_{gal}), 5.19 (1H, dt, H-2_{gal}), 5.35 (1H, dd, $J_{2,3} = 10.7$ Hz, H-3_{gal}), 5.50 (1H, d, $J_{3,4} = 3.2$ Hz, H-4_{gal}), 5.77 (1H, dd, $J_{1,2} = 3.5$ Hz, $J_{1,P} = 7.7$ Hz, H-1_{gal}), 5.94 (1H, d, $J_{1,2} = 3.5$ Hz, H-1_{rib}), 5.95 (1H, d, H-D_{ura}), 7.95 (1H, d, $J_{C,D} = 8.1$ Hz, H-C_{ura}).

**Synthesis of uridine 5'-(6-azido-6-deoxy- α -D-galactopyranosyl) diphosphate
bistriethylammonium salt**

The previously synthesized uridine 5'-(2,3,4-tri-*O*-acetyl-6-azido-6-deoxy- α -D - galactopyranosyl) diphosphate bistriethylammonium salt (0.270 g, 0.29 mmol) was dissolved into methanol (20 mL) at 0 °C. An aqueous solution of NH_4HCO_3 (27 mL, 0.1 M) and Et_3N (1.1 mL, 26 eq) was added to the mixture and stirred for 23 h at 0 °C. After 23 h, the solution was kept at 0 °C, diluted with water (120 mL), and the pH was adjusted to 7.5 with Dowex 50W-X8 H^+ form resin. The resin was removed by filtration and the resin was washed with water (60 mL). After lyophilization of the filtrate, the product was purified by semi-preparative HPLC C18-column (linear gradient for 10 min from 95:5 water/MeCN to 60:40 water/MeCN with detection at 254 nm) to afford the final product as a clear white solid (130 mg, 79%). ^1H -NMR (500 MHz, D_2O) δ 1.24 (18H, t, $(\text{CH}_3\text{CH}_2)_3\text{NH}^+$), 3.17 (12H, q, $J = 7.3$ Hz, $(\text{CH}_3\text{CH}_2)_3\text{NH}^+$), 3.42 (1H, dd, $J_{5,6'} = 6.1$ Hz, H-6'gal), 3.53 (1H, dd, $J_{5,6} = 7.2$ Hz, $J_{6,6'} = 12.7$ Hz, H-6gal), 3.75 (1H, dt, H-2gal), 3.88 (1H, dd, $J_{2,3} = 10.3$ Hz, H-3gal), 3.96 (1H, d, $J_{3,4} = 3.2$ Hz, H-4gal), 4.17-4.26 (4H, m, H-4rib, H-5gal, H-5rib, H-5'rib), 4.32-4.36 (2H, m, H-2rib, H-3rib), 5.59 (1H, dd, $J_{1,2} = 3.6$ Hz, $J_{1,P} = 7.3$ Hz, H-1gal), 5.93-5.97 (2H, m, H-1rib, H-Dura), 7.94 (1H, d, $J_{C,D} = 8.2$ Hz, H-Cura). ^{31}P NMR (121.5 MHz, D_2O) δ -13.00 (1P, d, $J_{P,P} = 21.0$ Hz, P-rib), -11.26 (1P, d, $J_{P,P} = 20.7$ Hz, P-gal).

¹H NMR

